U.S. Department
of Transportation
National Highway
Traffic Safety
Administration

www.nhtsa.gov

DOT HS 811 231 December 2009

Measuring Distraction Potential of Operating In-Vehicle Devices

DISCLAIMER

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT HS 811 231		
4. Title and Subtitle Measuring Distraction Potential of Operating In-Vehicle Devices		5. Report Date December 2009 6. Performing Organization Code NHTSA/NVS-312
7. Author(s) Thomas A. Ranney, G. H. Scott Baldwin, and Scott M. Vasko, Transportation Research Center Inc.; Elizabeth N. Mazzae, National Highway Traffic Safety Administration		8. Performing Organization Report No.
9. Performing Organization Name and Address National Highway Traffic Safety Administration Vehicle Research and Test Center P.O. Box 37 East Liberty, OH 43319		10. Work Unit No. (TRAIS) 11. Contract or Grant No.
12. Sponsoring Agency Name and Address National Highway Traffic Safety Administration 1200 New Jersey Avenue SE. Washington, DC 20590		13. Type of Report and Period Covered 14. Sponsoring Agency Code

15. Supplementary Notes

The authors wish to thank Mike Perel, Frank Barickman, Ed Parmer, Heath Albrecht, Adam Andrella, Cecilia Black, Guogang Xu, and Kamel Salaani for their contributions to this work.

Three experiments were conducted to explore the feasibility of adapting existing protocols to assess in-vehicle information systems (IVIS) in production vehicles. Two low-fidelity driving simulators were used: the Lane Change Task (LCT) and the STISIM-Drive combined with the Peripheral Detection Task (PDT). The Rating Scale Mental Effort (RSME) workload rating scale and FaceLab eye tracker were also used. Experiment 1 combined simulator driving with laboratory tasks (visual search and short-term memory scanning), which allowed secondary task load to be systematically varied. Metrics sensitive to changes in visual load included LCT Mean Deviation and the following STISIM/PDT metrics: Car-Following Delay, Standard Deviation of Lane Position (SDLP), Steering Entropy, PDT Mean Response Time, and Proportion of Correct PDT responses. Among objective metrics, only PDT Mean Response Time was sensitive to changes in cognitive load associated with the (auditory/vocal) memory-scanning task. Experiment 2 used real-world secondary tasks performed with a factory-installed navigation system. Secondary tasks differed by input modality (manual vs. voice) and task complexity (destination entry vs. selecting previous destinations). STISIM/PDT metrics, including SDLP, Steering Entropy, and Proportion of Correct PDT Responses were sensitive to task differences, as was the LCT Mean Deviation. STISIM/PDT metrics were more sensitive than LCT metrics to differences in both experiments. The RSME subjective rating scale was sensitive to most differences, while eye position data was not sufficiently reliable to allow computation of eye glance-based metrics. Experiment 3 used an established test-track protocol to determine whether measures obtained in a real driving situation exhibited greater sensitivity to potential distraction effects than those obtained in the simulation laboratory. Several laboratory simulator measures were more sensitive to secondary task load differences than corresponding test track measures. We concluded that the STISIM/PDT test venue offers sufficient sensitivity for development of a portable test of IVIS distraction potential in production vehicles for visual/manual tasks; improved sensitivity is needed for assess effects of cognitive distraction. We identified technical problems and questions about test validity to be addressed in subsequent developmental work.

17. Key Words	18. Distribution Statement
Driver performance, distraction, cognitive distraction, navigation system, in-vehicle information systems (IVIS)	Document is available to the public from the National Technical Information Service www.ntis.gov

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	94	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

PREFACE

The research described in this report was conducted in the 2006-2007 timeframe. Following this research, NHTSA undertook an additional effort in 2008-2009 to further develop the test and related metrics for the purpose of assessing the distraction potential of In-Vehicle Information Systems (IVIS). NHTSA anticipates release of a final report for the follow-on work in 2010. "References made to "future" or "additional" research in the current report should be assumed to refer to the 2008-2009 work.

EXECUTIVE SUMMARY

The measurement of distraction has been the focus of several large-scale projects undertaken by consortia of researchers, government agencies, and automotive manufacturers. These include the recently completed European project HASTE (Human machine interface And the Safety of Traffic in Europe), the Driver Workload Metrics (DWM) Consortium of the Collision Avoidance Metrics Partnership (CAMP), and the German Advanced Driver Attention Metrics (ADAM) program. The goal of these projects was to develop methodologies and guidelines for assessing the extent to which in-vehicle information systems (IVIS) interfere with driving.

Much of this work has been directed at evaluating pre-production versions of IVIS. As a result, very little consideration has been given to adapting protocols/metrics to assess IVIS that are already available in production vehicles. The National Highway Traffic Safety Administration (NHTSA) anticipated this need and undertook this project to explore the feasibility of adapting one or more existing protocols for this purpose. The work was conducted by researchers at NHTSA's Vehicle Research and Test Center (VRTC).

The first objective of this project was to select the most promising protocols/metrics. We considered protocols/metrics that had demonstrated sensitivity for detecting driving performance degradation associated with in-vehicle secondary tasks with either visual/manual or cognitive interfaces. Testing in production vehicles required selecting tests that could be administered without intrusion or vehicle damage caused by instrumentation. Accordingly, we only considered tests that could be administered without requiring significant vehicle modification. Additional criteria included: (1) the ease of implementation and administration, (2) the test protocol's state-of-development, including extent of use and documentation, (3) the level of training and staffing required, and (4) the availability and interpretability of data. Finally, objective measures were given preference over subjective measures.

Using these criteria, we selected two low-fidelity driving simulators as our primary test venues. These included the ADAM Lane Change Task (LCT) and the Systems Technology Inc. (STI) low-cost driving simulator (STISIM-Drive). The LCT was developed as a standalone desktop driving simulator with predefined scenarios and performance measures; it requires drivers to execute 18 lane changes in 3 minutes using information obtained from signs appearing in the scenario. The LCT combines vehicle control performance, object detection, and response speed into a single summary performance measure. The STISIM is a low-fidelity driving simulator. It is more generic than the LCT in that the researcher has control of the scenario events and performance measures. Based on CAMP study recommendations, we combined the Peripheral Detection Task (PDT) with the STISIM to provide an object-event detection component. We adapted scenarios developed by CAMP, which involved car following with occasional oncoming traffic. We used Rating Scale Mental Effort (RSME) workload rating scale, and the Seeing Machines faceLab eye tracking system with both primary test venues.

We conducted three experiments to evaluate the metrics associated with the two test venues. Experiments 1 and 2 were conducted in a simulator laboratory and required participants to drive both the STISIM/PDT combination and the LCT simulator. In Experiment 3 drivers performed secondary tasks while driving on a closed-course test track. The first two experiments assessed the metrics' sensitivity for detecting known and hypothesized differences between different secondary tasks. Experiment 1 used variations of laboratory tasks, which had been used in the aforementioned large scale projects. Although not realistic, these tasks offered the significant benefit of allowing secondary task load to be systematically varied. Metrics that differentiated between secondary task loads were rated higher than those that did not. Most metrics were sensitive to changes in visual/manual load associated with visual search tasks, but among objective metrics, only the PDT Mean Response Time was sensitive to changes in cognitive load associated with an auditory/vocal task. STISIM metrics, including Standard Deviation of Lane Position (SDLP), Steering Entropy, PDT Mean Response Time, and Proportion of Correct PDT Responses were the most sensitive objective metrics. LCT Mean Deviation was slightly less sensitive to the manipulations. The RSME subjective rating scale was sensitive to most differences.

Experiment 2 used real-world secondary tasks performed with a factory-installed navigation system. Secondary tasks differed in terms of input modality (manual vs. voice) and task complexity (destination entry versus selecting previous destinations). None of the metrics revealed differences between input modalities. This was due to the fact that the hands-free operation of this particular factory-installed navigation system applied only to a small subset of actions necessary to the complete the tasks. Both input modes required visual monitoring of the display and manual manipulation of the keyboard. The finding was interpreted not as a weakness of any metric, but rather as evidence that the voice interface in this particular IVIS was not substantively different from the visual/manual interface and thus did not significantly reduce distraction effects. Most metrics detected the difference between the simpler previous destination selection task, which consisted primarily of searching lists, and the more complex destination entry tasks, which required keyboard entry. Three STISIM measures (SDLP, Steering Entropy, and PDT Proportion Correct) were more sensitive to the differences between tasks than LCT summary measures. The RSME subjective rating scale was sensitive to most differences.

When the results from Experiments 1 and 2 were considered together, the driving performance metrics associated with the STISIM/PDT combination were shown to have slightly greater overall sensitivity than the two summary metrics that represent LCT performance. The subjective workload rating (RSME) was among the most sensitive measures and was consistent across the two simulator venues. Eye position data were considered to be of insufficient quality for computing metrics based on glance characteristics. In addition to the modestly greater sensitivity associated with the STISIM/PDT metrics relative to the LCT metrics, several practical considerations contributed to our decision to select the STISIM/PDT combination for further development. Specifically, the STISIM/PDT offers the flexibility necessary for exploring means to improve the sensitivity of the metrics to distraction effects that are primarily cognitive. In contrast, the LCT test is fixed. The breadth of STISIM/PDT measurement capabilities is also consistent with the general consensus that multiple measures are necessary to fully characterize distraction effects. Finally, the STISIM/PDT metrics were more amenable to comparison on the test track than the LCT metrics.

Experiment 3 was conducted to compare the sensitivity of measures obtained in the laboratory with that of an established test track protocol. We wanted to ensure that the simulators' inability to represent the full complexity of the driving task did not detract from the simulator metrics' sensitivity. The test-track protocol had been demonstrated to be sensitive to distraction effects for a variety of in-vehicle secondary tasks, including visual/manual and auditory/vocal tasks. Participants performed a subset of the same secondary tasks used in Experiments 1 and 2 while driving an instrumented vehicle on a closed test track with some traffic present. The similarity among patterns of workload ratings between laboratory and test track experiments implies that the simulator plus secondary task experience closely matched the test track plus secondary task experience. Beyond that, there were unexpected differences in sensitivity between the simulator and test track metrics. Specifically, for several measures, the laboratory simulator measures were more sensitive to secondary task load differences than the corresponding test track measures. Because the tasks in question had been used in previous experiments and the metrics in those studies had demonstrated sensitivity to the differences between levels, we attributed the differences in sensitivity to changes in our test track protocol plus inherent differences in the test environments. Specifically, the laboratory environment provided better control of test conditions, particularly visibility, and less measurement error than the test track. Participants' responses to differences in risk perception may also have contributed to the differences between lab and test track results. Nevertheless, with respect to the main objective of the third experiment, the results revealed no shortcomings of the simulator test venue. The limited fidelity of the simulator did not reduce the sensitivity of the simulator-based metrics for detecting the targeted differences between task conditions.

Based on the foregoing, we concluded that the STISIM/PDT test venue offers sufficient sensitivity and flexibility for continuing the development of a portable test of IVIS distraction potential in production vehicles. We identified several technical problems, including the need for greater sensitivity for measuring cognitive distraction and the need to improve the quality of the eye tracking data. Additional developmental work to address these needs, followed by an assessment of the modified test using a wider variety of production vehicles and real IVIS tasks, particularly those involving voice-based interfaces would be beneficial. Finally, the differential sensitivity between the simulator and test track venues must be reconciled before this test will provide information that can meaningfully be tied to safety. A more complete validation focused on determining the cause of the reduced sensitivity observed among test track measures would be beneficial.

1.0 INTRODUCTION

1.1 Background

The measurement of distraction has been the focus of several large-scale projects undertaken by consortia of researchers, government agencies, and automotive manufacturers. These include the recently completed European project HASTE (Human machine interface And the Safety of Traffic in Europe) (Carsten & Brookhuis, 2005a), the Driver Workload Metrics (DWM) Consortium of the Collision Avoidance Metrics Partnership (CAMP) (Angell et al., 2005) and the German Advanced Driver Attention Metrics (ADAM) program (Mattes, 2003). The goal of these projects has been to develop methodologies and guidelines for assessing the extent to which in-vehicle information systems (IVIS) interfere with driving.

The HASTE program was undertaken by eight European partners and Canada. Numerous experiments were conducted across Europe and Canada using a variety of test venues. One major finding was that the effects of cognitive distraction differ considerably from those of visual distraction. Secondary tasks that were mostly visual led to decrements in steering and lateral vehicle control. In contrast, secondary tasks that were mostly cognitive led to decrements in longitudinal vehicle control, particularly car following (Carsten et al., 2005a). HASTE researchers found differences between the testing venues. Specifically, they found that driving was degraded more on real roads than in simulators when drivers performed the same secondary tasks. They speculated that this discrepancy was due to the relatively limited fidelity of existing simulators. However, emphasizing the efficiency and reproducibility of the assessment environment provided by driving simulators over the realism of real-road driving, they concluded that an assessment regime that uses a reasonably advanced driving simulator with scenarios that require rural road driving, can provide meaningful and potentially reliable results (Carsten et al., 2005a; Carsten et al., 2005b). They also concluded that between four and six behavioral parameters would be needed to evaluate any system offered for assessment.

The Driver Workload Metrics (DWM) project was conducted by the CAMP consortium, which included researchers from Ford, GM, Nissan and Toyota. Their focus was on selecting driving performance metrics obtained in an experimental context that can be used to predict the safety implications of distraction in real driving. They conducted experiments in three test venues, including laboratory, test track and on-road driving. Four categories of driving performance metrics were identified as having direct implications for safety. These included driver eye glance patterns, lateral vehicle control, longitudinal vehicle control, and object-and-event detection. The researchers also identified a number of potential surrogates, which included laboratory measures, ratings and analytical methods thought to have predictive value with respect to the above-mentioned performance measures. They performed a series of analyses to determine which of their performance metrics discriminated driving with a secondary task from driving alone. They also determined which metrics discriminated high from low workload secondary tasks. The majority of metrics that passed one or both of these tests were eye glance measures. In addition, they found that measures generally discriminated high from low workload tasks much better for visual/manual than for auditory/vocal secondary tasks. Visual/manual tasks affected driving performance more than auditory/vocal tasks.

One significant conclusion of the CAMP project was that the interference to driving caused by in-vehicle secondary tasks was multidimensional and no single metric could measure all effects.

In agreement with the HASTE results, CAMP researchers found that visual/manual secondary tasks exhibited different performance profiles than auditory/vocal tasks. They concluded that eye glance data contain important information for assessing the distraction effects of both auditory/vocal and visual/manual tasks. Based on the secondary tasks they used, they concluded that cognitive distraction plays a much smaller role than visual distraction. Finally, because they found different patterns of degradation between the laboratory and on-road driving test venues, they concluded that the laboratory results alone were not sufficient to fully characterize the distraction potential associated with their secondary tasks.

The ADAM project has focused on the development of a lane change task (LCT). This task requires drivers to respond to a sequence of lane change assignments while performing secondary tasks (Mattes, 2003). The summary measure derived from the LCT has been shown to be sensitive to different types of secondary tasks and is being promoted as a standardized measure of distraction potential.

These projects were ambitious attempts to select driving performance metrics with some known relationship to on-road safety. However, as they progressed it became clear that it is virtually impossible to use experimental results to predict real-world risks associated with different secondary tasks. Thus, while the metrics identified in these studies may be very helpful for assessing the relative potential for distraction associated with in-vehicle systems during their development, the ultimate safety effects of new in-vehicle technologies cannot be known until the technologies are used in real-world driving, and data pertaining to drivers' willingness to engage in the secondary tasks are obtained.

1.2 Application to Production Vehicles

Much of the existing work has been directed at the need to evaluate pre-production versions of IVIS. As a result, very little consideration has been given to adapting protocols or measures to assess IVIS that are already available in production vehicles. The National Highway Traffic Safety Administration (NHTSA) anticipated a need to evaluate the distraction potential of technologies in production vehicles and sought to adapt one or more existing protocols for this purpose. To address this anticipated need, NHTSA undertook this project, which was conducted by researchers at NHTSA's Vehicle Research and Test Center (VRTC).

The first objective of this project was to select the most promising of the existing protocols or metrics that were suitable for our purposes. To be considered, protocols/metrics must first have demonstrated sensitivity for detecting interference associated with secondary tasks with either visual/manual or cognitive demands. The use of production vehicles requires that data be obtained without intrusion or damage to the vehicle caused by instrumentation. Accordingly, the next criterion for evaluating existing tests was that the test protocol could be implemented and data obtained without requiring vehicle modification. Additional criteria included: (1) the ease of implementation and administration, (2) the test protocol's state-of-development, including extent of use and documentation, (3) the level of training and staffing required, and (4) the availability and interpretability of data. Finally, objective measures were given preference over subjective measures.

Using these criteria, we evaluated materials from ongoing or recently-completed programs, including HASTE, ADAM, CAMP, and AIDE. In addition, we consulted with the research staff of Transport Canada, who were directly involved in several of these programs. Initially, we eliminated metrics requiring the use of instrumented vehicles for two reasons: first, we wanted a relatively portable test, not one requiring a closed course; second, we concluded that obtaining steering-based vehicle control metrics would necessitate unacceptable modifications to test vehicles. We selected two low-fidelity driving simulators as our primary test venues. These included the ADAM Lane Change Task (LCT) and the Systems Technology Inc. (STI) driving simulator (STISIM-Drive).

- The LCT was developed as a desktop simulator; it requires drivers to execute 18 lane changes in 3 minutes, using information obtained from signs appearing in the scenario. It has a single performance measure, which has been shown to be sensitive to the effects of both visual and cognitive distraction (Burns, Trbovich, McCurdie, & Harbluk, 2005; Mattes, 2003). The LCT is being developed as a draft ISO standard (ISO/TC 22, 2004).
- The STISIM is a low-fidelity driving simulator. It is more generic than the LCT in that the researcher has control of the scenario events and performance measures. We combined the STISIM with the Peripheral Detection Task (PDT), which has been used in numerous studies to measure changes in drivers' ability to detect targets reflected on the windshield (Harms & Patten, 2003b). We adapted scenarios developed and employed by CAMP, which involved steady-state car following with occasional oncoming traffic.

We combined the PDT with STISIM, as was done in the CAMP study for two reasons: first, CAMP researchers found the STISIM/PDT combination to be more promising than the STISIM car-following task alone; and second, the basic car-following task used in the STISIM did not have an object/event detection component. In contrast, the LCT has evolved as a complete, standalone test, including both vehicle control and discrete visual target detection components embedded in predefined scenarios. Adding the PDT to the LCT was considered to be unrealistic because it would add a second, potentially conflicting target-detection component. This conflict would adversely affect the LCT summary performance measure, rendering the results not comparable to LCT results more generally. PDT measures would be similarly confounded in such a combination. For this reason, we did not consider modifying the LCT. The comparison thus matches a fairly well established test (LCT), which combines vehicle control and target detection components, and provides a single summary measure versus a more loosely-defined framework in which a car-following task implemented on the STISIM is combined with a well-established detection task. The STISIM/PDT combination provides multiple performance measures, which is more consistent with the above-discussed consensus concerning the need for multiple measures to characterize distraction potential. The use of multiple measures raises the technical problem of how to establish weights and combine them into an overall assessment of distraction potential. This challenge does not exist for the LCT.

We used Rating Scale Mental Effort (RSME) workload rating scale, and the Seeing Machines faceLab eye tracking system (Victor, Harbluk, & Engstrom, 2005) with both test venues.

1.3 Research Objectives

The overall research objective was to evaluate the two selected test venues to determine which performance measures are most sensitive for detecting the interference caused by two categories of secondary tasks. The first category of secondary tasks included three calibration/reference tasks used in the HASTE and ADAM projects. These tasks were adapted from laboratory tasks. They were more abstract than real-world secondary tasks, but offered the significant benefit of allowing the secondary task load to be systematically varied. The second category consisted of real-world secondary tasks performed using an in-vehicle navigation system in a production vehicle. The specific tasks differed in terms of input modality (manual vs. voice) and task complexity. The tests using these categories of secondary tasks were incorporated into Experiments 1 and 2, respectively.

CAMP and HASTE researchers both found differences between laboratory and driving test venues and recommended that future metric evaluation studies include a driving component. More specifically, based on concerns expressed in the HASTE study (Carsten & Brookhuis, 2005a), we wanted to determine whether the simulators' limited fidelity adversely affected the sensitivity of the metrics. For this purpose, we used a VRTC test-track protocol, which has been demonstrated to be sensitive to distraction effects of various in-vehicle secondary tasks (Ranney, Harbluk, & Noy, 2005; Ranney, Mazzae, Baldwin, & Salaani, 2007). Participants performed secondary tasks while driving an instrumented vehicle on a closed test track with some traffic present. This experiment, therefore, served as a partial validation of the simulator studies, with particular emphasis on the effects of the simulators' reduced fidelity on the sensitivity of the corresponding metrics. Modifications necessary to instrument vehicles for test-track use were not acceptable for use with vehicles selected for testing, which would likely be borrowed or leased. Thus, the test track protocol was included as a focused validation, not as a candidate test venue.

1.4 Study Overview

This research study consisted of three experiments, each of which utilized a dual-task paradigm, in which a primary driving task (vehicle control plus object and event detection) was performed concurrently with a secondary task (specified interaction with an IVIS). The first two experiments used the two stationary test venues (LCT and STISIM/PDT), while the third experiment was a test track study. Two categories of secondary tasks were used, including a set of abstract reference or calibration tasks with known differences in information-processing load (Experiments 1 and 3), and a set of destination-entry tasks performed using the navigation system of a 2004 Acura TL (Experiments 2 and 3). All secondary tasks had two levels, which were either known (Experiment 1) or hypothesized (Experiment 2) to differ in their demand.

Experiment 3, which used selected tasks from both categories of secondary tasks, was conducted on the high-speed test track at the Transportation Research Center in East Liberty, Ohio; whereas, Experiments 1 and 2 were conducted in a stationary vehicle in leased lab space in Columbus, Ohio.

Our data analysis strategy was directed at answering specific questions about the metrics under consideration. Individual paired comparisons were conducted for each secondary task for each metric. The objective was to determine which metrics could differentiate between the two levels of each respective secondary task. Specifically, the analyses conducted in Experiment 1 were directed at determining which measures were sensitive to the differences in load manipulations of the calibration tasks. Metrics that differentiated between different loads were rated higher than those that did not. Similarly, the analyses conducted in Experiment 2 were directed at determining which measures were sensitive to differences in secondary task load associated with real world navigation system tasks. In Experiment 3, we examined the same differences, using a subset of the tasks from each of the first two experiments.

2.0 EXPERIMENT 1

2.1 Method

2.1.1 Participants

Twenty-six drivers (aged 25 to 50 years) participated in Experiment 1. Participants were recruited through advertisements placed in local newspapers and screened to ensure that they were active drivers with a valid driver's license and a minimum of 7,000 miles driven per year. Preference was given to participants who had experience using a wireless phone while driving. Data for Experiment 1 were collected during April and May of 2006.

2.1.2 Laboratory

Experiment 1 was conducted in a laboratory at The Ohio State University Center for Automotive Research (OSU CAR) located in the metropolitan area of Columbus, Ohio. The laboratory space consisted of a 12 foot by 36 foot room with no windows.

2.1.3 Apparatus

Components of the fixed based simulator included a production test vehicle (2004 Acura TL), an Intel Pentium 4 computer, a ceiling-mounted digital projector (1024 x 768) positioned over the vehicle, and a forward projection screen (8 feet x 8 feet), which was located approximately 12 feet in front of the driver's seated position. A touch screen was installed inside the vehicle and connected to a separate computer, used to generate stimuli for secondary tasks. A keypad was installed to record secondary task inputs. The touch screen and keypad are shown in Figure 1.

Figure 1. Acura Interior with Response Keypad

Sensors that recorded steering, accelerator and brake inputs were attached temporarily to the test vehicle. Specifically, a steering wheel overlay (see Figure 2) was developed to allow drivers to sit inside the Acura while operating the two driving simulators. This steering wheel provided steering inputs to the driving simulators, allowing the simulators to run without the vehicle being turned on. A vacuum pump was used to extend the range of brake pedal deflection with the vehicle off. Additional details of the technical modifications made to adapt the simulators for use with production vehicles are presented in Appendix A.

Figure 2. Steering Wheel Overlay

The Subject Vehicle (SV) MicroDAS data acquisition system (Barickman & Goodman, 1999) for Experiment 1 was configured to collect hand wheel position, brake and throttle inputs, and participant responses to the PDT. In addition, both the LCT and STISIM simulation computers collected data for their respective performance measures. The primary data channels for Experiment 1 are displayed in Table 1.

Table 1. Subject Vehicle Data Collection Channels for Experiments 1 and 2

Data Channel	Description	Units	Resolution
Vehicle Speed	STISIM	km/h	1 km/h
Range	Distance to the LV, STISIM	m	.5 m
Range-Rate	Relative velocity between the SV and the LV, STISIM	m/s	.1 m/s
Lateral Position	Lateral position of the SV in reference to the simulated lanes, STISIM and LCT	cm	2 cm
Hand Wheel Position	Angular position of the steering wheel (0 degrees = straight)	deg	.1 deg
UTC Time	Time of day	HH:MM:SS	1 s
Event Task	PDT button press	0 or 1	$1/30^{th}$ s

We modified the LCT for use in a stationary vehicle using an overhead projector and large screen. We used the steering wheel overlay described above to obtain driver steering inputs to the LCT. No accelerator inputs were required for the LCT as speed was held constant.

A Seeing Machines FaceLAB eye tracking system was used to record head pose and gaze. Head pose uses three parameters to define position and three parameters to define orientation.

FaceLAB outputs gaze rays for each eye. Each ray has an origin at the center of the respective eye and vectors pointing toward the object being looked at.

Gaze is represented as pitch and yaw angles. The pitch and yaw angles are transformed into a direction vector. Dual gaze is converted into a single gaze vector. The system used two stereo cameras mounted on the dashboard and was relatively unobtrusive. To assist the system in tracking facial features, participants applied five stickers to their faces during system calibration. The simulator plus secondary task setup is shown in Figure 3.

Figure 3. Simulator with Secondary Task (Experiment 1)

2.1.4 Procedure

Each participant completed one session, lasting approximately four hours.
Upon arrival, the participant was asked to read and sign the Participant Information Summary (Appendix B), thereby giving informed consent to participate in the study. No individuals declined to participate.

The participant was escorted to the experimental vehicle and given an overview of the vehicle controls and displays, including adjusting the seat and steering wheel. Next, the participant was given instructions and practice for the driving task components, including the PDT. Instructions and practice were then given for the secondary tasks. This was followed by an explanation of the monetary performance incentive system and the Rating Scale Mental Effort (RSME) (Appendix C).

The participant was then asked to affix latex markers to his or her face for use in eye tracker calibration. During this procedure, the experimenter instructed the participant concerning head position and point of gaze. Eye tracker calibration was completed. The participant was then given an opportunity to ask questions about any aspect of the protocol. Data collection began following a break. The experimenter was in the back seat of the vehicle during the data collection.

Experiment 1 consisted of approximately 20 driving trials, 10 in each test venue. Each trial lasted approximately three minutes. After each trial, the experimenter asked the participant to complete the RSME and provided performance feedback. The experimenter then described the next trial and read secondary task instructions aloud. The participant was given a break after each block of 10 driving trials.

At the completion of data collection, the participant exited the vehicle and completed a simulator sickness questionnaire (Appendix D) to determine if rest was required before being allowed to drive home. The experimenter paid the participant a total of two amounts: (1) Base pay for participation, and (2) Performance incentive pay. The experimenter answered any questions and returned the participant to his or her personal vehicle.

2.1.5 Driving Tasks

Lane Change Task (LCT). Developed as a desktop simulation, the ADAM Lane Change Task (LCT) requires drivers to execute 18 lane changes in 3 minutes, using information obtained from signs appearing in the scenario. In our implementation, the speed was fixed at 60.0 kph so that all participants had to respond to the same number of lane-changes during each 3-minute driving trial. Drivers sat in the Acura TL and responded to the moving roadway image projected onto the large screen in front of the vehicle. They manipulated the steering wheel overlay to provide steering inputs to the simulation. Figure 4 shows the roadway scene as a driver approaches one set of signs. Each sign has three fields corresponding to the three lanes of the roadway. The arrow (left in the figure) indicates that the driver is to move from the present lane (center) to the left-most lane. Drivers were instructed to perform each lane change as soon as possible after seeing the sign and to maintain vehicle control inside the lanes both before and after each lane change. Task performance was defined as the total deviation of the driver's path from the path associated with a normative model.

Figure 4. Lane Change Task Display

STISIM. A car-following paradigm modeled after that used by Brookhuis and colleagues (Brookhuis, Waard, & Mulder, 1994), was programmed into the scenario run on the STI simulator (see Figure 5). This task required participants to maintain a constant following distance behind a lead vehicle, which changed speed according to a predefined sinusoidal waveform. When implemented on the STI simulator, participants were required to follow a simulated lead vehicle's speed changes on straight road segments. Drivers were given training and feedback about the range of following distances considered acceptable. However, as in previous studies, because of individual differences in comfort associated with close following distances, a narrow range of following distances was not enforced. During the experiment, participants received feedback and monetary incentives based on their ability to maintain a consistent and relatively close following distance. An auditory warning system was also used to encourage drivers to maintain a fairly close following distance. When drivers exceeded a pre-defined criterion, an audible tone sounded once every five seconds until the driver returned to an acceptable following distance. The car-following task was always presented together with the peripheral detection task described below.

Figure 5. STISIM Visual Display

Peripheral Detection Task. The Peripheral Detection Task (PDT) has been used in numerous studies to measure changes in drivers' ability to detect targets reflected on the windshield (Harms & Patten, 2003a). The PDT consisted of an array (3 x 20 cm) of 23 high-intensity (12,000 mcd) LEDs (as illustrated in Figure 6) positioned on the dashboard and shielded from direct view of the driver. LED activation appeared as a reflection in the windshield located at positions with eccentricities ranging between approximately 5-25º to the left of the driver's line of sight and 2-4º above the dashboard.

This dashboard-mounted version with windshield reflections requires participants to detect targets at fixed locations. At intervals that varied randomly between 3 and 5 seconds, one of the 23 LEDs was illuminated. Each LED activation lasted 1.5 s, unless terminated by the driver's response. Drivers responded as quickly as possible by pressing a micro switch attached to their left index finger. Valid responses were defined as responses recorded between 200 ms and 2000 ms following LED activation. Response times and proportion of targets detected were computed for each secondary task trial.

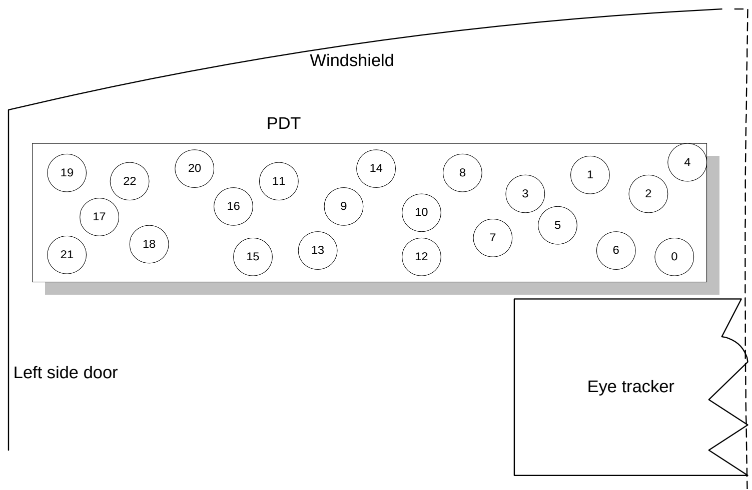
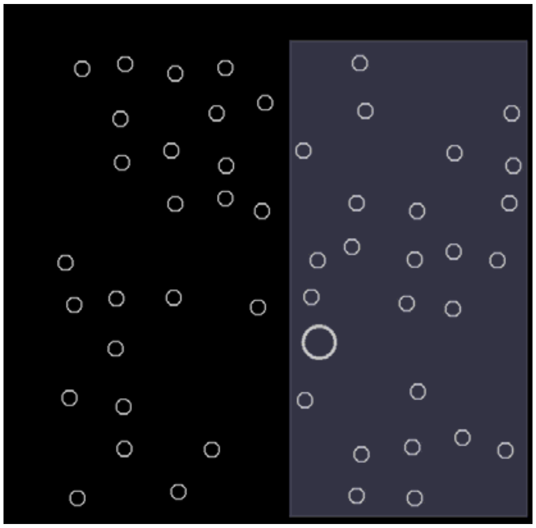


Figure 6. Layout of PDT LEDs

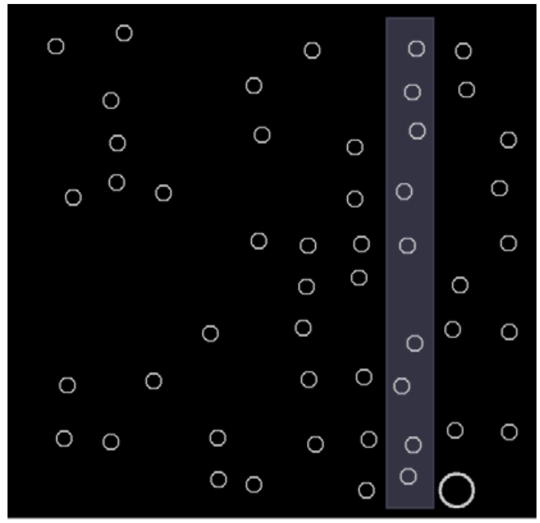
2.1.6 <u>Secondary Tasks</u>

Secondary tasks used in this experiment included the Circles task (ADAM), the Arrows task (HASTE) and the Sternberg memory scanning test (ADAM). These tasks are described below.

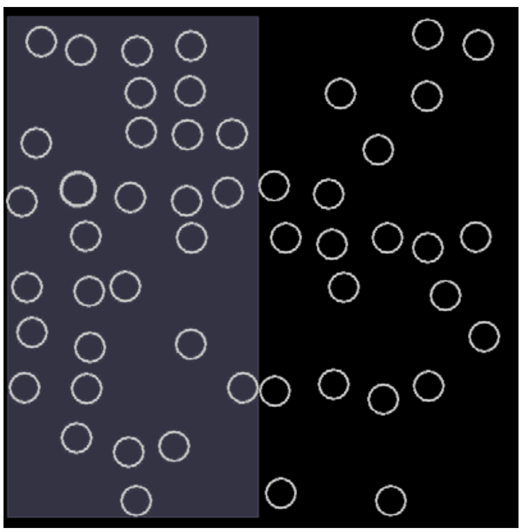
<u>Circles Task.</u> This self-paced visual search task required participants to view a sequence of arrays of circles displayed on a computer screen, located inside the test vehicle. Each array contained a single larger circle among an array of smaller circles. Each array also included a darker region (vertical band), which could be moved across the screen with keystrokes (left or right arrows). Participants were required to move the vertical bar until it covered the target circle. Visual task difficulty was manipulated by varying the size of the larger target circle relative to the array of smaller distractor circles. Manual (motor) difficulty was manipulated by varying the size of vertical band. Four conditions were used for testing, including all combinations of 2 levels of visual difficulty and 2 levels of motor difficulty. Specifically, in the easy visual condition, the distractor circles were 50% of the target circle size, while in the more difficult visual condition this value was 83%. In the easy motor condition, the vertical band covered half the visual display so that at most one keystroke was required to identify the target location. In the difficult motor condition, the vertical band had 10 locations, which required multiple keystrokes to identify the target location. Examples of each condition are shown in Figure 7.



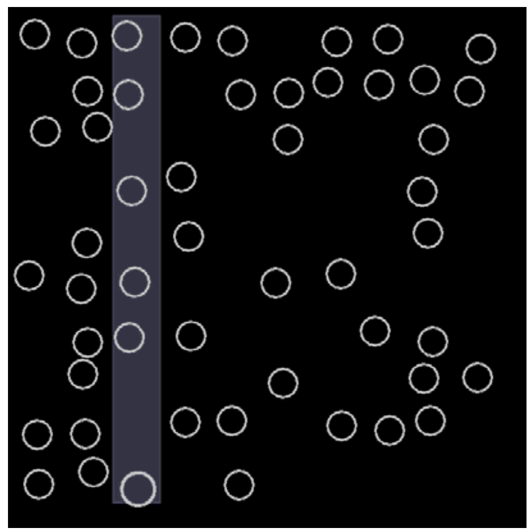
Condition D75M02 (50%, 2 bands)



Condition D75M10 (50%, 10 bands)



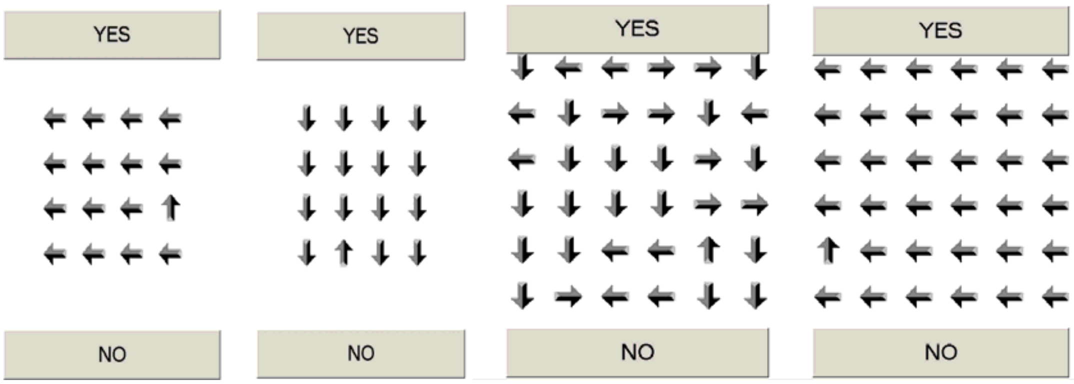
Condition D125M02 (83%, 2 bands)



Condition D125M10 (83%, 10 bands)

Figure 7. Circles Task Stimuli

Arrows Task. Adapted from the HASTE program, this externally-paced visual search task was designed to require primarily visual processing and minimal cognitive processing (Jamson & Merat, 2005). The task requires participants to view a sequence of arrays of arrows displayed on a touch-screen LCD mounted in the vehicle. Participants searched for a single upward pointing target among a matrix of distractor arrows. The target arrow was present on 50% of the trials. Task difficulty was defined by the number of arrows on the display. Matrices were either 4 x 4 or 6 x 6. Within each difficulty there were different types of patterns, including those in which all arrows except the target pointed in one direction and those in which arrows had any orientation. On each driving trial, participants would view a series of 36 matrices, all with the same number of arrows present. A new array was presented every 5 seconds. Participants responded by pressing the 'Yes' or 'No' button on the touch screen, reflecting their decision about the presence of the target arrow. Figure 8 shows examples of the two conditions.



Difficulty level 1 (easier condition) is shown on the left with two examples of the 4 x 4 matrix, and difficulty level 3 (more difficult condition) is shown on the right with two examples of the 6 x 6 matrix.

Figure 8. Arrows Task Stimuli

Sternberg (Number Memory) Task. The Sternberg (memory scanning) task was adopted by ADAM as a cognitive reference task for the Lane Change Task. On each trial, the participant is presented with a set of digits (3 or 7 digits depending on the task difficulty), referred to as the memory set. The participant is then presented a single digit and must determine (yes/no) whether this single digit is among those in the memory set. All digits were presented using voice recordings; therefore, there were no visual stimuli for this task. Two difficulty levels were used; they differed only in the number of memory set digits (3 or 7). Memory-set digits were presented a rate of 1 per second. This was followed by a 15-second period of silence, during which the participant had to remember the memory set. A single target digit was then presented and participants then had 5 seconds to respond aloud. Individual trials thus lasted between 24 and 28 seconds depending on the memory set size. Participants completed 6 trials during each 3-minute driving trial.

Participants completed twenty 3-minute drives, ten in each test venue. The drives included one while performing each of the 8 secondary task conditions plus 2 baselines, with no secondary tasks, in each testing venue.

2.1.7 Monetary Incentives

Participants were given a base pay of $20 per hour, plus monetary incentives to motivate acceptable performance. Monetary rewards were awarded based on experimenter ratings as shown in Table 2. Incentive amounts were defined to establish priorities among the three task components. For example, to emphasize driving as the highest priority, the LCT or car-following task was associated with the highest monetary values.

Table 2. Experiment 1 Incentive Amounts per Trial

		Performance			
Test Venue	**Task**	**Priority**	**Good**	**Acceptable**	**Poor**
Lane Change Task	Lane Change Task	1	$0.60	$0.30	$0.0
	Secondary Task	2	$0.40	$0.20	$0.0
	Total		$1.00	$0.50	$0.0
STISIM	Car Following	1	$0.60	$0.30	$0.0
	Secondary Task	2	$0.40	$0.20	$0.0
	Light-Detection	3	$0.20	$0.10	$0.0
	Total		$1.20	$0.60	$0.0

During each session, participants in Experiment 1 completed approximately 20 trials: on each of the 10 LCT trials, the participant had the opportunity to earn $1.00, for a total of $10.00; on each of the 10 STISIM trials, the participant had the opportunity to earn $1.20, for a total of $12.00. Thus, for good performance, each participant could earn an additional $22.00.

2.1.8 Data Reduction

Data from the STISIM trials were reduced to compute the following driving performance measures:

Coherence. Coherence is a measure of squared correlation, which reflects the degree to which the following vehicle is able to match the periodicity of the lead vehicle speed signal. Coherence is used both as a measure of car-following performance and as a test of whether the associated measure of phase shift (car-following delay) is interpretable. Coherence requires a car-following paradigm in which the lead vehicle speed changes can be represented as a combination of sine waves. Technical details describing the computation of coherence are presented in Appendix F.

Phase Shift (Delay in Car Following). This measure represents the response lag in car following. Its interpretation is similar to that of discrete response time measures in that longer delay values reflect poorer performance than shorter values. When coherence is relatively high (e.g., ≥ 0.80), the driver is adequately following the lead vehicle's speed changes, which implies that the associated measures are meaningful. When coherence values are low, the estimates of phase shift (delay) are considered suspect. We therefore included phase shift values in our analysis only for trials for which coherence was greater than 0.8. Less than 5% of the data were eliminated for this problem.

Headway. While driving, participants were instructed to maintain a constant following distance during all trials. Our previous work (Ranney et al., 2005), as well as that of Brookhuis (Brookhuis, De Vries, & De Waard, 1991), has shown that drivers have considerable difficulty maintaining a prescribed following distance. We therefore allowed drivers to select their own following distance and encouraged them to maintain that distance. However, we have seen that despite instructions, some drivers increased their following distances while performing secondary tasks. This measure has been interpreted as reflecting compensation for increased demands during secondary task performance, relative to baseline driving.

Standard Deviation of Lane Position (SDLP). This measure reflects the variability of lateral position over the entire data collection interval. It has been widely used as a measure of driving performance and has been shown to be sensitive to impairment due to fatigue, alcohol, drugs and distraction.

Steering Entropy. Developed by Boer (Boer, 2000), steering entropy measures the error associated with loaded conditions (secondary task present) relative to a designated baseline run. The measure is based on autocorrelation and represents the frequency and extent of high-frequency corrections following periods when the driver's visual attention is diverted from the roadway.

PDT Mean Response Time. Drivers responded to approximately 20 targets during each driving trial. Responses recorded between 0.2 and 2.0 seconds following the target activation were considered correct responses. Mean response time is computed for the correctly detected targets on each trial.

PDT Proportion Correct. This measure represents the proportion of PDT targets detected correctly on a given trial.

Head Position X Std. We intended to use eye gaze measures provided by the FaceLAB eye tracking system. However, preliminary examination of the data quality ratings revealed that the data were both laden with a relatively high percentage of erroneous data and highly unstable with respect to the positional representation of the same gaze location. Head position measures were found to be more stable than eye position measures. They represent an estimate of which way the head is pointing. We considered this measure to be roughly comparable in quality to information provided by manually-reduced eye position information. Following Victor et al. (Victor et al., 2005), we computed the standard deviation of X (horizontal) position, which represents the variability in side to side movement during each trial.

RSME Workload Ratings. This scale (Appendix C) represents the participants' ratings of the subjective difficulty associated with each combination of primary and secondary task.

Data from the LCT were reduced to compute the following performance measures:

Mean Deviation. This measure represents the average instantaneous deviation between a participant's lateral position and a standardized normative model over the entire 3-minute drive during which the participant completed 18 lane change events.

Mean Deviation (Individual Baselines). This measure is computed in the same way as the above measure, with the exception that the participant's own baseline run is used, rather than the normative model.

Head Position X Std. Same as above.

RSME Workload Ratings. Same as above.

2.2 Results

We used Proc Mixed of SAS (Version 9.1.3) to compute an analysis of variance (ANOVA) for each dependent measure. Secondary task was the independent variable. It had the following nine levels:

1. Arrows D1 – Easier: 4 x 4 matrices
2. Arrows D3 – Harder: 6 x 6 matrices
3. Baseline – No secondary task
4. Circles V1M1 – Easy visual discrimination, easy motor response
5. Circles V1M2 – Easy visual discrimination, harder motor response
6. Circles V2M1 – Harder visual discrimination, easy motor response
7. Circles V2M2 – Harder visual discrimination, harder motor response
8. Sternberg 3 – Easier memory task
9. Sternberg 7– Harder memory task

Our data analysis was directed at answering specific questions about the metrics under consideration. Specifically, in Experiment 1, we were interested in determining which of our candidate metrics were sensitive to the load manipulations within the calibration tasks. We therefore identified the following planned comparisons:

1. Arrows: D1 vs. D3 (Visual/manual load)
2. Circles: V1 vs. V2 (Visual load, collapsed across Manual conditions)
3. Circles: M1 vs. M2 (Manual load, collapsed across Visual conditions)
4. Sternberg : 3 vs. 7 (Cognitive load)

Because we specified planned comparisons, we did not interpret omnibus *F* values. Separate *F* tests were computed for each planned comparison for each performance measure. Probability values were adjusted for familywise error by using Hochberg's step-up method (Westfall, Tobias, Rom, Wolfinger, & Hochberg, 2003). Adjusted *p* values of less than .05 are considered to be statistically significant. Adjusted *p* values between .05 and .10 were considered marginal and discussed where applicable. A summary of the results of the planned comparisons with adjusted *p* values is presented in Table 3 for the STISIM/PDT measures and Table 4 for the LCT measures. Means for each performance measure by secondary task condition are presented in Figure 9 (pp. 19-20).

Table 3. Summary of Planned Comparison STISIM Results Experiment 1

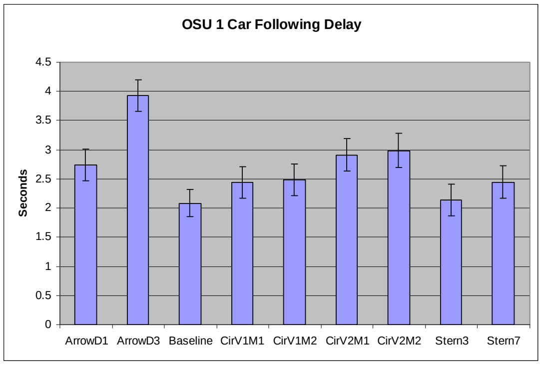
Task	Comparison		Delay	Cohere	Mean Hdwy	SDLP	Steer. Entropy	PDT MRT	PDT P Corr	Head X Std	RSME
Arrow	D1 vs. D3		.0001*	.0034*	(.17)	< .0001*	.0003*	< .0001*	< .0001	.001	< .0001
Circle	V1 vs. V2		.03*	(.36)	.02*	< .0001*	< .0001*	.0006*	< .0001	< .0001	< .0001
Circle	M1 vs. M2		(.77)	(.53)	(.97)	.01*	.07+	(.59)	(.40)	(.76)	(.16)
Sternberg	3 vs. 7		(.36)	(.53)	(.97)	(.72)	(.14)	.01	(.25)	(.55)	< .0001

* Statistically significant difference ($p < .05$)
+ Marginally not significant ($.05 < p < .10$)
Parentheses denote differences that were not statistically significant

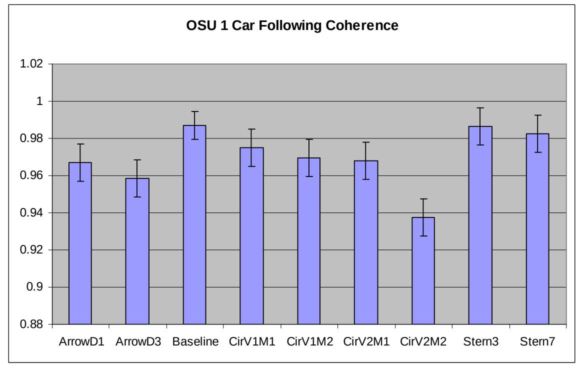
Table 4. Summary of Planned Comparison LCT Results Experiment 1

Task	Comparison		Mean Deviation	Mean Deviation (Indiv. BL)	Head X Std	RSME
Arrow	D1 vs. D3		.0005*	.004*	.0002*	< .0001*
Circle	V1 vs. V2		.014*	(.12)	<.0001*	.0015*
Circle	M1 vs. M2		(.65)	(.67)	(.90)	(.41)
Sternberg	3 vs. 7		(.73)	(.67)	(.90)	.0001*

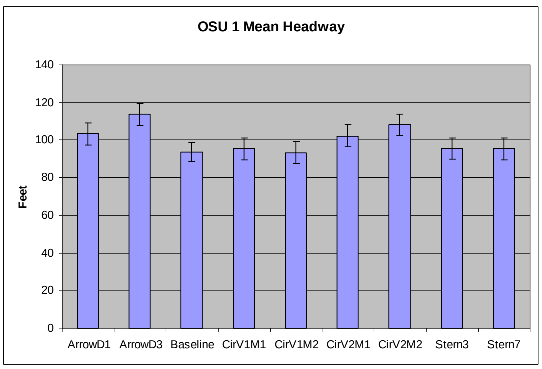
* Statistically significant difference ($p < .05$)
+ Marginally not significant ($.05 < p < .10$)
Parentheses denote differences that were not statistically significant



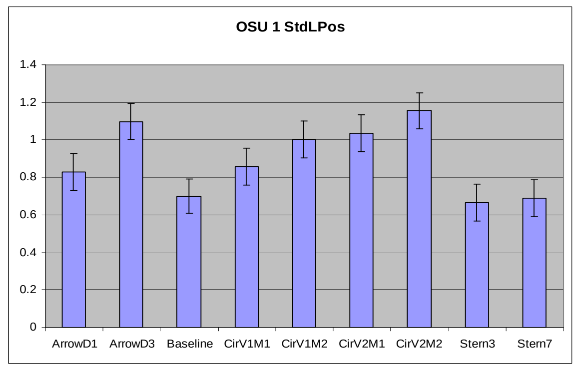
(a) STISIM Car Following Delay



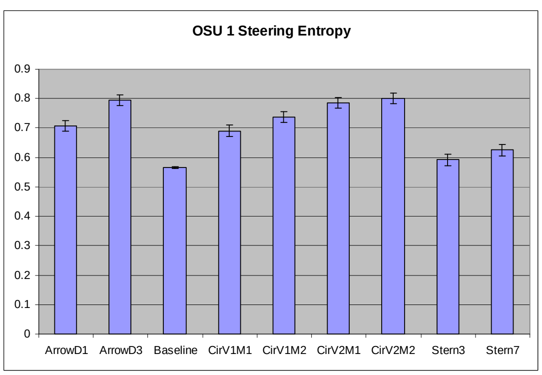
(b) STISIM Car Following Coherence



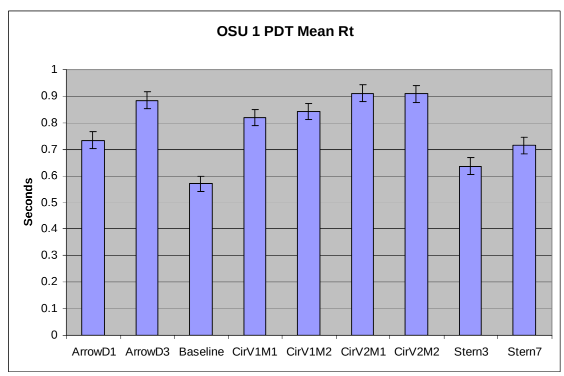
(c) STISIM Mean Headway



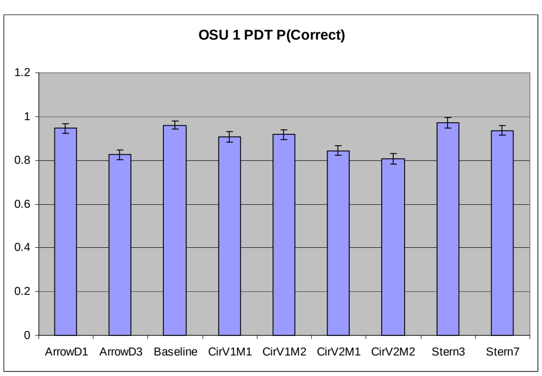
(d) STISIM SDLP



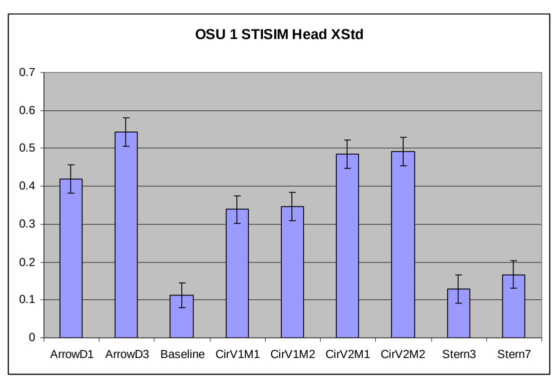
(e) STISIM Steering Entropy



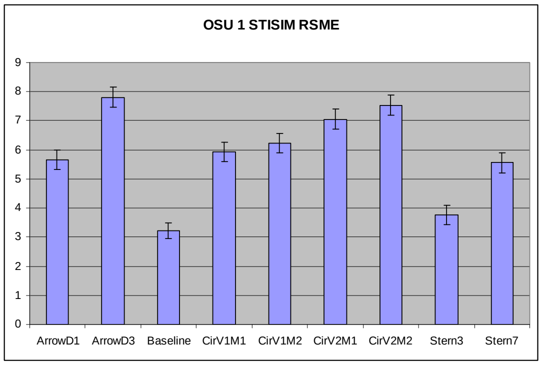
(f) STISIM PDT Mean RT



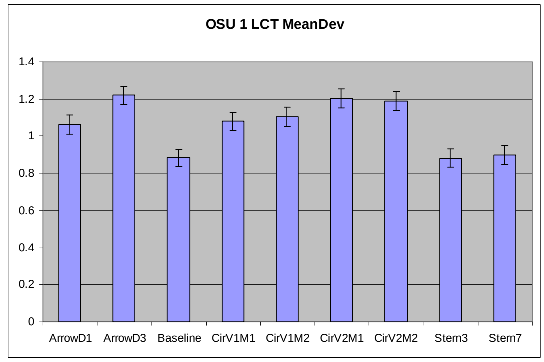
(g) STISIM PDT Proportion Correct



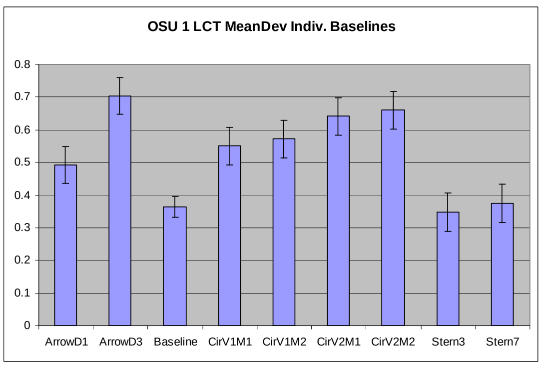
(h) STISIM Head X Std



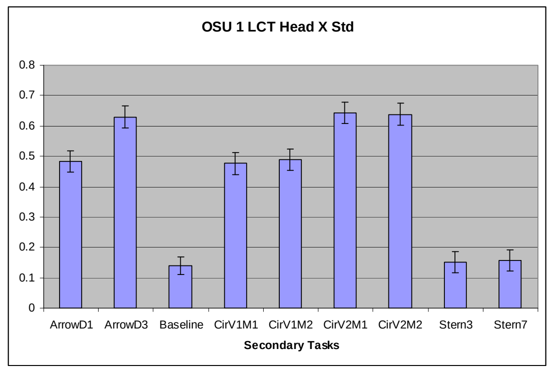
(i) STISIM RSME



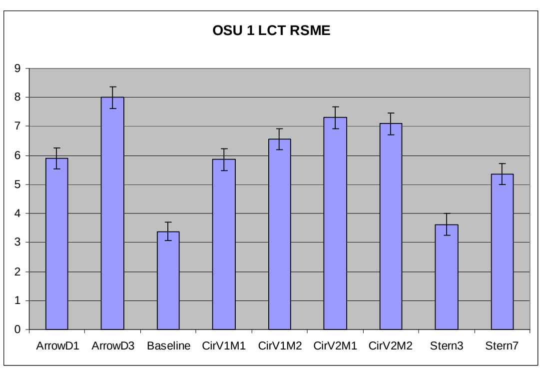
(j) LCT Mean Deviation



(k) LCT Mean Deviation (Individual Baselines)



(l) LCT Head X Std



(m) LCT RSME

Figure 9. Mean Values (± Standard Error) for Each Metric by Secondary Task Condition – Experiment 1

2.3 Discussion

Although statistical analyses were focused on the four specific comparisons identified above, it is noteworthy, as shown in Figure 9, that Baseline values for the metrics were generally seen to be extreme values. (STISIM Mean Headway is an exception.) This confirms that the metrics were generally sensitive to the loads associated with the secondary tasks used in this experiment.

The first planned comparison (PC 1) focused on differences between Arrows Task conditions, representing differences in visual/manual load. Results for PC 1 are shown in row 1 of Table 3 and Table 4. The results indicate that most performance measures were sensitive to the differences between the two Arrow Task conditions. This task was perhaps the most demanding of the four tasks in that it was externally paced and not under the driver's control.

Differences between Circles Task conditions were examined in the second and third planned comparisons (PC 2 and PC 3). PC 2 focused on differences in visual task demand while PC 3 focused on differences in manual task demand. Most performance measures were sensitive to the changes in visual demand associated with the Circles task (PC 2 row 2 in Table 3 and Table 4); however, only one measure (standard deviation of lane position - SDLP) was sensitive to the difference in the manual component of the Circles Task (PC 3, row 3). Steering entropy was marginally sensitive to this effect.

The fourth planned comparison (PC 4) compared performance degradation due to differences between levels of the Sternberg memory scanning task (Stern 3 vs. Stern 7). The difference is in cognitive load. As shown (row 4 in Table 3 and Table 4), most performance measures were not sensitive to differences between the two auditory/vocal task conditions, reflecting a lack of sensitivity to cognitive distraction. Only PDT Mean RT and RSME revealed differences between these two conditions. The driving performance metrics thus were very sensitive to differences in visual load between secondary task conditions, but not particularly sensitive to differences due to manual or cognitive load.

Measures derived from the car-following task (coherence, delay, mean headway) were less sensitive than other measures to the differences between the specified task conditions in this experiment. However, they appeared to show relatively large effects of secondary tasks generally.

Overall, SDLP and PDT Mean RT, both associated with STISIM, were more sensitive than the other objective measures. RSME in both test venues was also sensitive to load differences between secondary task conditions.

3.0 EXPERIMENT 2

3.1 Method

3.1.1 Participants

Twenty-seven drivers (aged 25 to 50 years) participated in Experiment 2. Participants were recruited through advertisements placed in local newspapers and screened to ensure that they were active drivers with a valid driver's license and a minimum of 7,000 miles driven per year. Preference was given to participants who had experience using a wireless phone while driving. Data for Experiment 2 were collected during July and August of 2006.

3.1.2 Laboratory

Experiment 2 was conducted using the same laboratory space as Experiment 1 in The Ohio State University Center for Automotive Research (OSU CAR).

Figure 10. Acura TL Dashboard with Navigation System

3.1.3 Apparatus

We used the same fixed-base simulator components and eye-tracking system as in Experiment 1. We used the navigation system of the Acura TL, which consisted of a touch screen, a video display, and a set of input buttons (see Figure 10). An external antenna was positioned outside the laboratory to maintain system communication with GPS satellites so that the system could identify vehicle location. User inputs to the system were made with a combination of voice commands, touch screen entries, and button presses.

We modified the PDT response button for this study to remove the potential conflict between the button press required for this task and those required to activate the voice command system of the Acura TL navigation system. Specifically, we mounted the response button and transmitter on the left side of the steering wheel in a way that did not interfere with either the voice command button press or vehicle steering.

3.1.4 Procedure

Each participant completed one session, lasting approximately four hours. Upon arrival, the participant was asked to read and sign the Participant Information Summary (Appendix B), thereby giving informed consent to participate in the study. No individuals declined to participate.

The participant was escorted to the experimental vehicle and given an overview of the vehicle controls and displays, including adjusting the seat and steering wheel. Next, the participant was given instructions and practice for the driving task components, including the PDT. Instructions and practice were then given for the secondary tasks. This was followed by an explanation of the monetary performance incentive system (see Section 3.1.7) and the Rating Scale for Mental Effort (RSME).

The participant was then asked to affix latex markers to his or her face for eye tracker calibration. During this procedure, the experimenter instructed the participant concerning head position and point of gaze. Eye tracker calibration was completed. The participant was then given an opportunity to ask questions about any aspect of the protocol. Data collection began following a break. The experimenter was in the back seat of the vehicle during the data collection.

Experiment 2 consisted of 16 driving trials, 8 in each test venue. Each trial lasted approximately 3 minutes. After each trial, the experimenter asked the participant to complete the RSME and provided performance feedback. The experimenter then described the next trial and read secondary task instructions aloud. The participant was given a break after each block of 8 driving trials.

At the completion of data collection, the participant exited the vehicle and completed a simulator sickness questionnaire to determine if rest was required before being allowed to drive home. The experimenter paid the participant a total of two amounts: (1) Base pay for participation, and (2) Performance incentive pay. The experimenter answered any questions and returned the participant to his or her personal vehicle.

3.1.5 Driving Tasks

Drivers performed the same driving tasks as in Experiment 1, including the Lane Change Task (LCT) and a car-following task plus peripheral detection task (PDT) in the STISIM simulator.

3.1.6 Secondary Tasks

Secondary tasks used in this experiment included functions of the Acura TL navigation system. Specific tasks included destination entry by address, destination entry by places of interest (POI), and selecting a previous destination. Each task was performed with both the visual/manual and voice interfaces.

Destination Entry by Address. This self-paced task required participants to use the Acura TL navigation system to enter a sequence of addresses. Addresses were presented one at a time on a touch screen located inside the test vehicle to the right of the navigation system (see Figure 11). For each destination, the participant performed the sequence of operations listed below. After each address was entered, the participant touched the touch screen. This recorded the time to complete the address entry and displayed the next destination. The sequence of operations included:

Select Enter destination by Address button,
Press Street button
Enter street name via keyboard
Select street name from a scrolled list of streets
Enter street number via keyboard

All steps except keyboard entry could be accomplished via voice commands. Voice commands required pressing a button on the steering wheel to alert the system to expect a voice command and then speaking keywords recognized by the navigation system.

Figure 11. Acura Navigation System and Stimulus Touch Screen

<u>Destination Entry by Place of Interest</u>. This self-paced task required participants to enter a sequence of places using the Acura TL navigation system. Drivers performed this task repeatedly. They used the touch screen to view the next destination and to register the end of each entry. Manual place entry involved the following sequence of operations:

Select Enter destination by Places button
Select Find Place by Category button
Choose category, and subcategory if applicable
Enter place name via keyboard
Select Place from scroll list

<u>Selecting Previous Destination</u>. This self-paced task required participants to use the Acura TL navigation system to select a sequence of destinations that had previously been entered into the system. Drivers performed this task repeatedly during each 3-minute drive, obtaining new places via the touch screen, as above. The manual version of selecting previous destinations included:

Select Enter Destination by Previous Destination button
Select destination from scrolled list

The secondary tasks used in Experiment 2 were selected to be consistent with those used in experimental work by researchers at Transport Canada. Figure 12 shows a driver in the STISIM driving simulator performing car following while engaged in a secondary task using the Acura TL navigation system.

Figure 12. Navigation System plus Touch Screen Inside Test Vehicle

3.1.7 <u>Monetary Incentives</u>

Participants earned a base pay of $20 per hour, plus incentives based on their performance. Incentive amounts were identical to those used in Experiment 1 (see Table 2). Incentive amounts were selected to establish the car-following or the LCT as most important, followed by the in-vehicle secondary task and light-detection task, respectively. Participants in Experiment 2 completed 16 trials: on each of the 8 LCT trials, the participant had the opportunity to earn $1.00, for a total of $8.00; on each of the 8 STISIM trials, the participant had the opportunity to earn $1.20, for a total of $9.60. Thus, for good performance, each participant could earn an additional $17.60. For completing all 16 trials, participants were given an additional $4.40 bonus such that each person could earn up to $22.00 in addition to their base pay, as in Experiment 1.

3.1.8 <u>Data Reduction</u>

Data were reduced to obtain the same measures that were used in Experiment 1.

3.2 Results

We used Proc Mixed of SAS (Version 9.1.3) to compute an analysis of variance (ANOVA) for each dependent measure. Secondary task was the independent variable. It had the following seven levels:

1. Baseline – no secondary task
2. Manual Destination Entry by Address
3. Manual Destination Entry by Place

4. Manual Selection of Previous Destination
5. Voice Destination Entry by Address
6. Voice Destination Entry by Place
7. Voice Selection of Previous Destination

The analyses focused on 9 planned comparisons (PCs), which are summarized in Table 5. Tests were developed based on two hypotheses: (1) tasks performed using the voice interface would be less disruptive to driving than the same tasks performed with the visual/manual interface; and (2) Selecting a previously entered destination would be less disruptive to driving than either type of destination entry (i.e., by Address or Place of Interest) because it did not require keyboard use. PCs were based on these hypotheses. The first PC focused on the overall difference between interface conditions. PCs 2 and 3 examined the performance effects of different tasks performed using the visual/manual interface. PCs 4 and 5 examined the same differences for tasks performed using the voice interface. PCs 6-9 examined differences between task conditions using data from both interface conditions.

Table 5. Planned Comparisons for Experiment 2

	Comparison	Conditions	Collapsed Across:
1	Visual/Manual vs.Voice	2,3,4 vs. 5,6,7	Task condition
2	Manual Address vs. Manual Previous Destination	2 vs. 4	None
3	Manual (Add + POI) vs. Manual Previous Destination	2,3 vs. 4	None
4	Voice Address vs. Voice Previous Destination	5 vs. 7	None
5	Voice (Add + POI) vs. Voice Previous Destination	5,6 vs. 7	None
6	Address vs. Previous Destination	2,5 vs. 4,7	Interface
7	Point of Interest vs. Previous Destination	3,6 vs. 4,7	Interface
8	Address + POI vs. Previous Destination	2,3,5,6 vs. 4,7	Interface
9	Address vs. Point of Interest	2,5 vs. 3,6	Interface

Separate *F* tests were computed for each planned comparison. Probability values were adjusted for familywise error by using Hochberg's step-up method (Westfall et al., 2003). Adjusted *p* values of less than .05 are considered to be statistically significant; however several tests with marginal results are considered noteworthy. A summary of the results of the planned comparisons with adjusted *p* values is presented in Table 6 for the STISIM/PDT metrics and Table 7 for the LCT metrics. Means for each performance measure by secondary task condition are presented in Figure 13 (pp. 30-31).

Table 6. Summary of Planned Comparison Results STISIM Measures Experiment 2

	Comparison		Delay	Cohere	Mean Hdwy	SDLP	Steer Entropy	PDT MRT	PDT P Corr	Head X Std	RSME
1	Visual/Manual vs. Voice		(.37)	(.30)	(.35)	(.52)	(.30)	(.18)	(.32)	(.13)	(.49)
2	Manual Address vs. Manual Previous Dest.		(.51)	(.30)	.08+	.05*	.04*	(.99)	.0007*	.03*	.003*
3	Manual(Add + POI) vs. Manual Prev. Dest.		(.20)	(.30)	.01*	.05*	.04*	(.18)	< .0001*	.02*	.001*
4	Voice Address vs. Voice Previous Dest.		(.20)	(.76)	.01*	< .0001*	.01*	(.18)	.0001*	(.14)	.002*
5	Voice (Add + POI) vs. Voice Prev. Dest.		(.37)	(.60)	.01*	.0001*	.01*	(.18)	<.0001*	(.14)	.0007*
6	Address vs. Previous Destination		(.20)	(.60)	.003*	< .0001*	.002*	(.33)	< .0001*	.02*	< .0001*
7	Point of Interest vs. Previous Destination		(.20)	(.20)	.003*	.006*	.02*	.04*	< .0001*	.03*	.0002*
8	Address + POI vs. Previous Destination		(.20)	(.30)	.001*	< .0001*	.002*	(.11)	< .0001*	.02*	< .0001*
9	Address vs. Point of Interest		(.91)	(.47)	(.98)	(.13)	(.30)	(.18)	(.87)	(.62)	(.68)

* Statistically significant difference ($p < .05$)

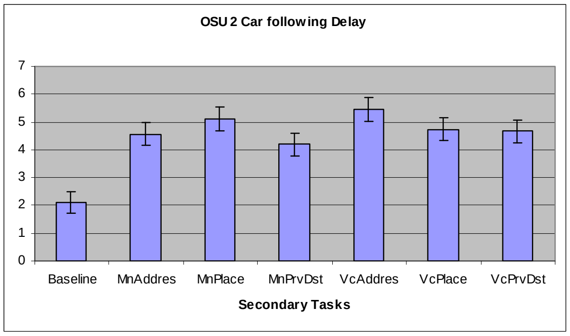
\+ Marginally not significant ($.05 < p < .10$)

Parentheses denote differences that were not statistically significant

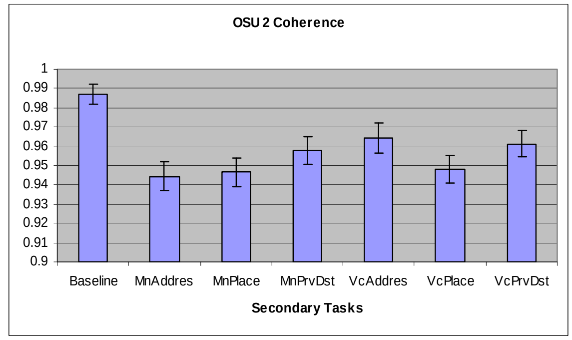
Table 7. Summary of Planned Comparison Results LCT Measures Experiment 2

	Comparison		Mean Deviation	Mean Deviation (Indiv. BL)	Head X Std	RSME
1	Visual/Manual vs. Voice		(.65)	(.64)	(.18)	(.53)
2	Manual Address vs. Manual Previous Dest.		.07+	.05*	.05*	.007*
3	Manual (Add + POI) vs. Manual Prev. Dest.		.07+	.051+	.06+	.002*
4	Voice Address vs. Voice Previous Dest.		.01	.007*	.005*	.03*
5	Voice (Add + POI) vs. Voice Prev. Dest.		.04	.01*	.007*	.02*
6	Address vs. Previous Destination		.005	.003*	.002*	.001*
7	Point of Interest vs. Previous Destination		.08+	.05*	.03*	.002*
8	Address + POI vs. Previous Destination		.01	.004*	.002*	.0004*
9	Address vs. Point of Interest		(.11)	(.16)	(.18)	(.84)

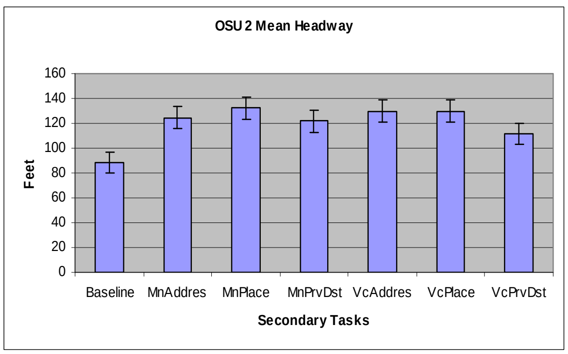
* Statistically significant difference ($p < .05$)
+ Marginally not significant ($.05 < p < .10$)
Parentheses denote differences that were not statistically significant



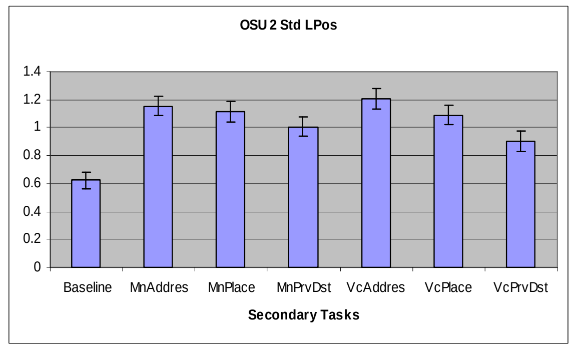
(a) STISIM Car Following Delay



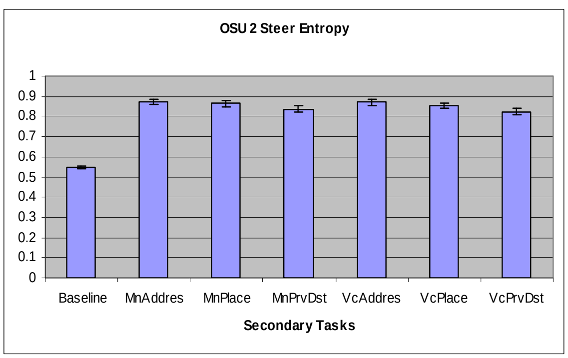
(b) STISIM Car Following Coherence



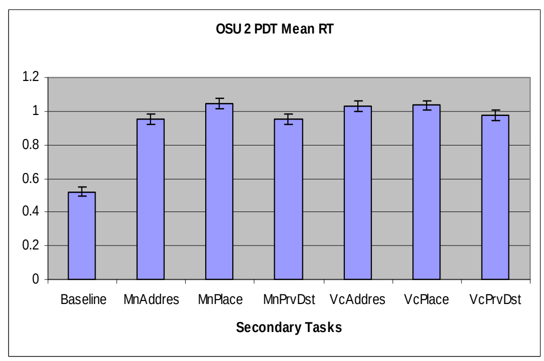
(c) STISIM Mean Headway



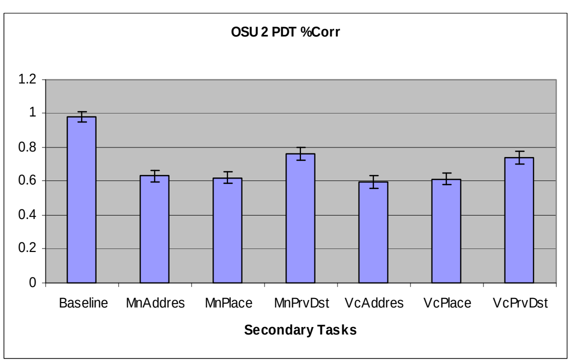
(d) STISIM SDLP



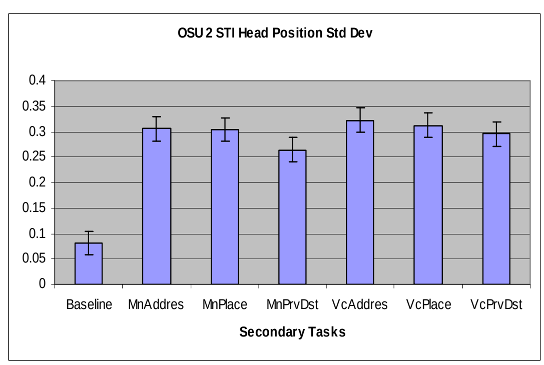
(e) STISIM Steering Entropy



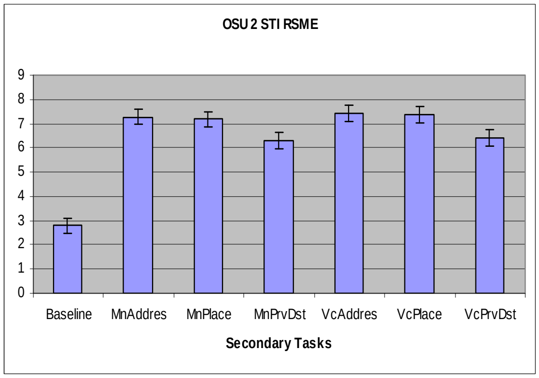
(f) STISIM PDT Mean RT



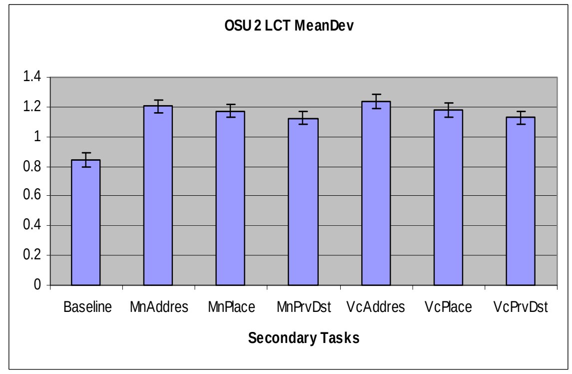
(g) STISIM PDT Proportion Correct



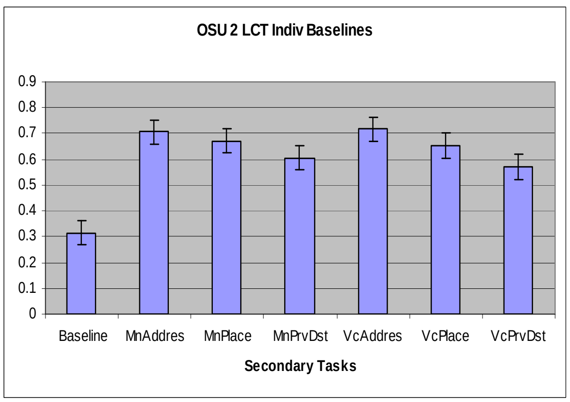
(h) STISIM Head X Std



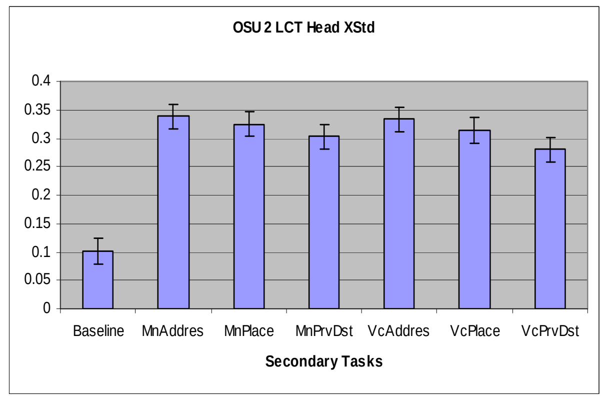
(i) STISIM RSME



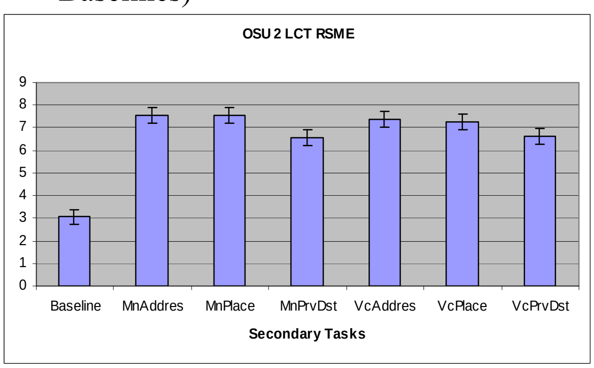
(j) LCT Mean Deviation



(k) LCT Mean Deviation (Individual Baselines)



(l) LCT Head X Std



(m) LCT RSME

Figure 13. Mean Values (± Standard Error) for Each Metric by Secondary Task Condition – Experiment 2

3.2.1 Effects of Interface on Secondary Task Performance

We analyzed video records of drivers' navigation system interactions to count the number of actions and record the time required to complete each entry. For our purposes, each action (button presses or voice commands) was counted. These data are summarized in Figure 14 and Figure 15.

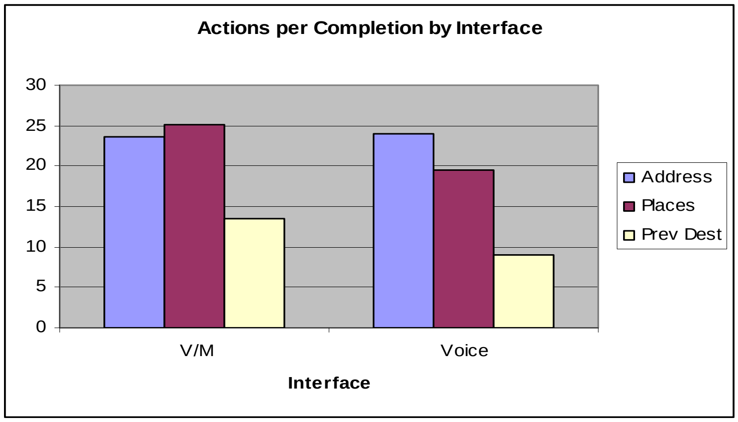


Figure 14. Driver Actions per Task Completion by Interface Condition (Experiment 2)

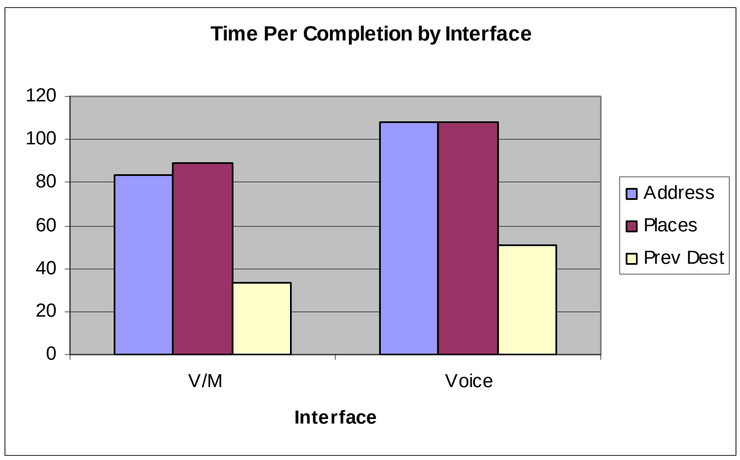


Figure 15. Secondary Task Completion Time by Interface Condition (Experiment 2)

3.3 <u>Discussion</u>

Inspection of Figure 13 reveals that all metrics exhibited relatively large differences between the baseline and most secondary task conditions. Although the observed differences were not tested statistically, they support the conclusion that the metrics' were generally sufficiently sensitive for detecting performance decrements due to secondary tasks loads used in this experiment. We now consider the nine planned statistical comparisons (PCs). Metrics derived from both simulators are considered.

<u>PC 1: Visual Manual vs. Voice</u>. This comparison used data from all 6 secondary task conditions; the three visual/manual conditions were compared with the three voice conditions. None of the measures revealed differences associated with PC 1, which supports the conclusion that there are no differences between the visual/manual and voice interfaces associated with the navigation system used in Experiment 2. In retrospect, this seems obvious in that there were only minor differences between the visual/manual and voice interface versions of the secondary tasks used in this experiment. Specifically, both interfaces required monitoring the visual interface and manual keyboard entry. Only the preliminary commands that involved selection of menu items or scrolling through lists could be performed with voice commands and these

required a button press to alert the navigation system to upcoming voice input. The voice interface used in this study was thus a hybrid interface and quite different from voice interfaces that allow hands-free interaction and have no associated visual display.

Figure 14 presents the number of discrete actions per completed destination by interface. Although these differences were not tested statistically, the voice interface was associated with a modest reduction in the number of discrete actions required for two of the three conditions, namely (destination entry by) places and previous destination; destination entry by address showed no such benefit. However, the results presented in Figure 15 lead one to question these apparent benefits. Specifically, when the time per completion is compared by task across interfaces, the voice interface condition was consistently associated with longer completion times. Drivers were thus more time efficient using the visual/manual interface.

The absence of differences associated with PC 1 highlights the difficulties associated with the use of real navigation system tasks in Experiment 2. In Experiment 1, secondary tasks were artificial but provided well-established differences in processing load. The secondary tasks used in Experiment 2 were realistic and accurately represent the types of tasks performed by drivers with navigation systems. However, we did not have an independent objective basis for determining the level of loading associated with these secondary tasks. This created problems interpreting the results of Experiment 2. However, we found the RSME workload rating scores to be helpful in addressing this problem. Because the RSME ratings in Experiment 1 closely followed the known differences between secondary task conditions, we decided to treat RSME findings as an independent assessment of whether or not to expect differences between conditions in Experiment 2. For example, the absence of RSME differences associated with PC 1 in Experiment 2 was consistent with the pattern of results for the objective measures, supporting the conclusion of no differences between navigation system interfaces. Therefore, in the absence of independent objective assessments of secondary task demands, metrics were considered successful if they provided results consistent with patterns of RSME differences.

<u>PC 2: Manual Address vs. Manual Previous Destination</u>. This comparison used data from the visual/manual conditions only to assess differences in demand between destination entry by address and selection of a previous destination. As noted above, these two conditions differed primarily in that the address condition required keyboard use while the previous destination required scrolling through lists, which was considered less demanding. Several of the STISIM/PDT metrics, including SDLP, Steer entropy, the percentage of correct PDT responses, the variation of head position in the X (side to side) direction (Head X Std) and RSME all were sensitive to this difference. Among the LCT metrics, the individual baseline mean deviation scores were statistically different, while the difference between means computed with the standard baseline score was marginally not significant.

<u>PC 3: Manual (Address + POI) vs. Manual Previous Destination</u>. This comparison used data from the visual/manual conditions to assess differences between the previous destination selection task and the two methods of destination entry (by address and by point of interest [POI]). This comparison is based on the hypothesis that the two destination entry tasks (address and POI) were similar in their level of demand/interference since both required some keyboard entry. With the exception of car-following measures (delay and coherence) and the PDT MRT, all STISIM/PDT measures successfully detected the difference between these conditions.

Among LCT measures, both mean deviation scores were marginally not sensitive to this difference.

PC 4: Voice Address vs. Voice Previous Destination. This comparison used data from the voice interface conditions to assess differences between the destination entry by address and selection of previous destination conditions. Relative to the corresponding comparison for manual data (PC 2), the metrics exhibited approximately equal sensitivity, with several exceptions. Car following headway means were different here, but not in the corresponding manual condition. The head-position metric (Head X Std) means were not different here. All LCT metrics successfully detected the differences between these conditions.

PC 5: Voice (Address + POI) vs. Voice Previous Destination. This test compared the two destination entry conditions versus the selecting previous destination condition using data from the voice interface condition only. It is the counterpart for PC 3, which made the same comparison using data form the visual/manual interface. Due to the absence of overall differences between visual/manual and voice interfaces, we expected the pattern of results to be similar to that of PC 3. We found this to be the case; five of the nine STISIM/PDT metrics detected the differences between these conditions. All LCT metrics were sensitive to this difference.

PC 6: Address (grouped) vs. Previous Destination (grouped). This comparison included data from both interface conditions, and following the finding in PC 1, we expected the pattern of differences to be similar to the findings of PC 2 and 4. Accordingly, six of nine STISIM/PDT metrics demonstrated differences. All LCT metrics were sensitive to this difference.

PC 7: Point of Interest (grouped) vs. Previous Destination. This comparison used data from both interface conditions. Seven of nine metrics demonstrated sensitivity for detecting this difference, including the PDT Mean Response time; this was the only difference detected by this metric in this experiment. Three of the four LCT metrics detected this difference; means for the standard mean deviation score were marginally not significant.

PC 8: Address + POI (grouped) vs. Previous Destination (grouped). In addition to combining data from both interface conditions, this test compared data from both destination entry conditions (address and POI) to the previous destination selection condition. The results were generally consistent with previous tests; six of nine measures revealed differences between the grouped conditions. All LCT metrics detected this difference.

PC 9: Address (grouped) vs. Point of Interest (grouped). This test compared the two methods of destination entry using data from both interface conditions. None of the metrics were sensitive to this difference, including RSME. This was true both for STISIM/PDT metrics and LCT metrics. The consistency of the findings supports the conclusion that this was not a real difference.

Based on the outcome of PC 1, we expected the results of PCs 2, 4 and 6 to be consistent in showing that any differences between these tasks would be independent of interface condition. All three comparisons considered differences between address destination entry and previous destination selection. PC 2 used visual/manual data only; PC 4 used the voice interface data; PC

6 used both sets of data. With minor exceptions, the three comparisons provided similar results; six of nine metrics demonstrated sensitivity in PC 6, versus five of nine for PC 2 and PC 4. The marginally greater sensitivity observed in PC 6 is likely due to the larger (combined) data set used for this comparison, which provided more statistical power, relative to the smaller data sets used in PC 2 and PC 4.

The results of PC 9 indicate that the destination entry by address task was not different from the destination entry by place (POI) task. None of the measures, including RSME, revealed differences between these conditions, using data collapsed across interface conditions. Both destination entry tasks required the combination of menu search, scrolling, and manual keyboard entry. In contrast, the select previous destination task required only scrolling through a list and no keyboard entry. It is for this reason that the comparisons based on combining the two destination entry tasks (PC 3, PC 5, and PC 8) provide essentially the same information as those that used only the destination entry by address data (PC 2, PC 4, and PC 6).

Considering the measures, it appears that the STISIM car-following measures (Delay and Coherence) were not useful in detecting differences between the secondary task conditions in this experiment. The same was true for PDT Mean Response Time, although this was likely due to our decision to move the response button location. Most consistent among the STISIM measures were standard deviation of lane position (SDLP), PDT proportion correct and steering entropy, followed by mean headway. Among LCT measures, mean deviation computed from individual baselines was more sensitive than the mean deviation computed with the normative model.

3.4 <u>Comparison of Results for Experiments 1 and 2</u>

To simplify interpretation of results from Experiments 1 and 2, we developed a three-point rating system based on the number of differences detected by each metric. Table 8 presents the criteria for these ratings. For example, metrics that detected 3 differences in Experiment 1 would be rated as highly sensitive, as would those that detected 6 or 7 differences in Experiment 2. Differences that were marginally not significant (.05 < p < >10) were weighted at half the value of a statistically significant difference based on the conclusion that these results reflect some level of sensitivity. Table 9 presents the ratings for each metric using the criteria presented in Table 8.

Table 8. Rating Categories for Experiments 1 and 2

Rating	Experiment 1	Experiment 2
High	3	6-7
Mid	2	4-5
Low	1	1-3

Table 9. Metric Rating Results Based on Results of Experiments 1 and 2

		Number of differences detected			Rating	
Measure	Test Venue	Expt. 1	Expt. 2	Total	Expt. 1	Expt. 2
Delay	STISIM	2	0	2	Mid	Low
Coherence	STISIM	1	0	1	Low	Low
Mean Headway	STISIM	1	6	7	Low	High
Std Lane Pos.	STISIM	3	7	10	High	High
Steering Entropy	STISIM	2.5	7	9.5	Mid	High
PDT Mean RT	STISIM	3	1	4	High	Low
PDT % Correct	STISIM	2	7	9	Mid	High
Head X Std	STISIM	2	5	7	Mid	Mid
RSME	STISIM	3	7	10	High	High
Mean Deviation	LCT	2	5.5	7.5	Mid	Mid
Mean Dev. (Ind. Bl)	LCT	1	6.5	7.5	Low	High
Head X Std	LCT	2	6.5	8.5	Mid	High
RSME	LCT	3	7	10	High	High
Note: Sums were computed using 1 for each statistically significant effects and 0.5 for each marginally non-significant result (.05 < p < >10).						

Standard deviation lane position (SDLP), obtained in the STISM, was the only objective measure rated High/High. RSME, a subjective assessment of workload was also rated High/High in both simulators. The following metrics were rated Mid/High or High/Mid:

- Steering entropy
- PDT proportion correct
- Head X Std (LCT)

Two of these metrics were associated with the STISIM/PDT test venue. It is unknown why the third metric, which was recorded in both test venues, was more sensitive in the LCT, however the visual demands of the driving task differed between the STISIM and LCT scenarios. We consider these to be the most promising measures based on the results of Experiments 1 and 2.

4.0 EXPERIMENT 3

4.1 Method

4.1.1 Participants

Twenty-nine drivers (aged 25 to 50 years) participated in Experiment 3. Participants were recruited through advertisements placed in local newspapers and screened to ensure that they were active drivers with a valid driver's license and a minimum of 7,000 miles driven per year. Preference was given to participants who had experience using a wireless phone while driving. Data for Experiment 3 were collected between November 2006 and February 2007.

4.1.2 Test Track

The experiment was conducted on the Transportation Research Center's (TRC) 7.5-mile oval test track, located in East Liberty, Ohio. The track consists of three 12-foot wide concrete lanes plus a fourth inner blacktop lane. Two straight segments, each approximately 2.0 miles long are separated by curved and banked segments, which are approximately 1.75 miles in length. Other traffic, including a mix of passenger vehicles and trucks, all traveling in the same direction, was present during data collection. The two experimental vehicles (described below) used the rightmost concrete lane. Occasionally, slower moving vehicles necessitated a lane change into the middle concrete lane; however, when this occurred, initiation of data collection was deferred until the lane change was completed. Similarly, stopped traffic was occasionally present in the inner blacktop lane. This created a temporary visual distraction, but did not otherwise interfere with the data collection. None of the trials was disrupted by slower or stopped traffic in the data collection lane. Data collection was suspended during inclement weather (e.g., when windshield wipers were required) and otherwise at the discretion of the experimenter, who monitored the speed and proximity of other traffic on the test track.

4.1.3 Apparatus

Two vehicles, including a lead vehicle (LV) and a subject vehicle (SV) were used to implement a car-following paradigm. Both vehicles were equipped with automatic transmissions, Micro Data Acquisition Systems (MicroDAS) (Barickman et al., 1999) and GPS receivers. GPS position readings were used to determine lane position and to derive vehicle speed. A Vorad radar device on each vehicle measured range (inter-vehicle spacing) and range rate to the other vehicle. The SV had a secondary brake for emergency activation by the experimenter accompanying the subject. The SV also had an event switch, which the experimenter used to mark the start and end of each data collection event in the data stream. The two data acquisition systems collected data independently at a 30-Hz sampling rate.

Subject Vehicle. The same Acura TL used in Experiments 1 and 2 was instrumented for this purpose. The SV MicroDAS was configured to collect vehicle speed, range, range-rate, lateral position, hand wheel position, GPS timing signals, and subject responses to the PDT. A video camera recorded the participants' button presses on the navigation system touch screen and surrounding control panel. An audio recorder captured the participants' voice inputs. The primary SV data collection channels are displayed in Table 10.

Table 10. Subject Vehicle Data Collection Channels for Experiment 3

Data Channel	Description	Units	Resolution
Vehicle Speed	Ground speed	km/h	1 km/h
Vorad Range	Distance to the LV	m	.5 m
Range-Rate	Relative velocity between the SV and the LV	m/s	.1 m/s
Lateral Position	Lateral position of the SV in reference to the center of the lane delineated by the painted edge markings	cm	2 cm
Lateral Velocity	SV Lateral velocity in reference to the painted edge markings	cm/s	2 cm/s
Road Curvature	Curvature of the upcoming roadway	m^{-1}	6.3e-10 m^{-1}
Offset Confidence	Reliability estimate of the lateral position	%	1 %
Road Curvature Confidence	Reliability estimate of the curvature data	%	1 %
Hand Wheel Position	Angular position of the steering wheel (0 degrees = straight)	deg	.1 deg
UTC Time	Time of day	HH:MM:SS	1 s
Pulse Per Second	GPS pulse per second signal used to synchronize data from both platforms	0 or 1	+/- 1 µs
Event Task	PDT button press	0 or 1	1/30th s

A Seeing Machines faceLAB eye tracking system was used to record head and eye movements. The system used two stereo cameras mounted on the dashboard and was relatively unobtrusive.

Lead Vehicle. The LV MicroDAS was configured to collect vehicle speed, tailway (distance to the SV), and GPS timing signals. The primary LV data collection channels are displayed in Table 11.

Table 11. Lead Vehicle Data Collection Channels for Experiment 3

Data Channel	Description	Units	Resolution
Vehicle Speed	Ground speed	km/h	1 km/h
Range	Distance to SV	m	.5 m
Range-Rate	Relative velocity between the LV and the SV	m/s	.1 m/s
UTC Time	Time of day	HH:MM:SS	1 s
Pulse Per Second	GPS pulse per second signal used to synchronize data from both platforms	0 or 1	+/- 1 µSec

The LV was equipped with a speed controller, which was created by interfacing a portable computer (486DX) with a servo controller running a basic proportional integral derivative control loop. The computer's data acquisition board generated analog signals that were sent to the servo controller. A user interface allowed the driver to select an input file and activate and deactivate the controller. The vehicle speed input, as measured by the LV's transmission speed sensor, provided feedback for the system.

For this experiment we used a trigonometric sine function with frequency of 0.03-Hz and extreme speed values of 50 and 65 mph. The associated acceleration and deceleration requirements were within limits of normal driving (i.e., < .4 G). Before the speed controller could be engaged, the vehicle had to be traveling at least 55 mph.

4.1.4 Procedure

Each participant completed one session, lasting approximately five hours. Upon arrival, the participant was asked to read the Participant Information Summary, which described the experiment and set forth the terms of participation. Participants also read a Confidential Information form (PP153) for visitors to the TRC proving ground, which describes TRC's policy for safeguarding proprietary information (Appendix E). After all questions were answered, the participant signed the documents, thereby giving informed consent to participate in the study. No individuals declined to participate.

The participant was escorted to the experimental vehicle and given an overview of the vehicle controls and displays, including adjusting the seat and steering wheel. Next, the participant was given test track guidelines, followed by instructions and practice for the driving task components, including car following and the PDT. Instructions and practice were then given for the secondary tasks. This was followed by an explanation of the monetary performance incentive system and the Rating Scale Mental Effort (RSME).

The participant was then asked to affix latex markers to his or her face for eye tracker calibration. This allowed the system to use facial features to help determine point of gaze and head position. During this procedure, the experimenter instructed the participant concerning head position and point of gaze. Eye tracker calibration was completed. The participant was then given an opportunity to ask questions about any aspect of the protocol. Data collection began following a break. The experimenter, who was seated in the back, was able to communicate directly with the LV driver (an experimental confederate) via two-way radio. Similarly, both the LV driver and the experimenter were able to communicate directly with the test track control tower. This provided notice of circumstances that would require them to stop or slow down.

Experiment 3 required drivers to complete 9 laps of the 7.5-mile test track, including training and practice. We collected data on the two 2.0-mile straight segments of each lap. Additional laps were included if a planned trial was aborted due either to equipment malfunction or traffic in the travel lane. Each trial required approximately 2.5 minutes. At the beginning of the data collection interval, the lead vehicle driver activated the LV speed controller which implemented the sine wave speed signal described above. The sine wave was deactivated at the end of the data collection interval. The experimenter instructed the participant when to begin and end each trial. Both vehicles stopped between trials on the rightmost blacktop lane. During the stops, the participant completed the RSME and was given performance feedback and new secondary task instructions. On the experimenter's signal, the LV driver accelerated and when traffic permitted, moved into the rightmost concrete lane and gradually increased speed to 55 mph.

Secondary task conditions (see Section 4.1.6) were separated into two blocks, including: (1) Reference/Calibration tasks, and (2) Navigation tasks. Baseline drives, with no secondary task, were included in each block. Practice trials were included at the beginning of each block. Practice began with car following alone. Drivers were encouraged to maintain a following distance of no more than 2 seconds, which represents a fairly conservative following distance for real-world driving on suburban freeways. Participants selected their own following distances; however, those who chose relatively long following distances were encouraged to adopt shorter following distances. Next, the PDT was added, followed by the secondary tasks. Drivers then stopped between trials, completed the RSME and were given performance feedback.

Occasionally, there was a vehicle stopped in the designated travel lane and the driver of the lead vehicle had to change lanes. The participant was instructed to change lanes whenever the lead vehicle did, as long as it was safe to do so. In this way, the LV driver ensured a safe path ahead on the test track. The LV driver would make decisions about changing lanes whenever possible before activating the speed signal at the beginning of the straightaway to minimize any potential effects of the lane change on the trial.

At the completion of data collection, the participant was paid two amounts: (1) Hourly base pay; and (2) Performance incentive pay. The experimenter answered any questions and returned the participant to his or her personal vehicle.

4.1.5 Driving Tasks

Car following. A car-following paradigm modeled after that used by Brookhuis and colleagues (Brookhuis et al., 1994), was used. This task required participants to maintain a constant following distance behind a lead vehicle, which changed speed according to a predefined sinusoidal waveform. When implemented on the TRC test track, participants were required to follow lead vehicle speed changes on each of the (2-mile) straight road segments. During training, drivers were given feedback about the range of following distances considered acceptable. To accommodate individual differences in comfort associated with close following distances, a narrow range of following distances was not enforced, during the experiment; however, participants received monetary incentives based on their ability to maintain a consistent and relatively close following distance.

Peripheral Detection Task (PDT). The Peripheral Detection Task (PDT), described in Section 3.1.5, was used in Experiment 3 (see Figure 16). The response button and transmitter were attached to the left side of the steering wheel as in Experiment 2.

Figure 16. In-Vehicle PDT Location with Single LED Activated

4.1.6 Secondary Tasks

Experiment 3 secondary tasks included 5 calibration/reference tasks (from Experiment 1) and 3 navigation system tasks (from Experiment 2) (see Table 12). There were also two baseline drives, in which drivers performed the primary task (car-following+PDT) alone. Thus, each participant completed 10 main test trials.

Table 12. Secondary In-Vehicle Task Descriptions for Experiment 3

Secondary task	Description	Levels	Stimulus presentation	Response mode
Circles Task	Designate location of larger circle among pattern of smaller circles	3	Visual	Physical manipulation of arrow keys
Sternberg Task	Identify designated digits among stream digits	2	Voice	Verbal response
Destination Entry by Address	Enter street address and city	1	Visual	Combination of voice and touch screen
Destination Entry by Places	Find specified place (e.g. restaurant, hotel)	1	Visual	Combination of voice and touch screen
Select Previous Destination	Search list of destinations previously entered	1	Visual	Combination of voice and touch screen

The Circles and Sternberg memory scanning tasks used in Experiment 3 were identical to those used in Experiment 1. Similarly, the navigation system tasks used in Experiment 3 were identical to the voice interface tasks used in Experiment 2.

4.1.7 Monetary Incentives

In addition to a base pay of $20 per hour, participants had the opportunity to earn a modest amount of additional money during the experiment. The actual amount of money awarded per trial was based on participants' performance in the three tasks shown in Table 13. Incentive amounts were intended to establish the car-following task as most important, followed by the in-vehicle secondary task and light-detection task, respectively.

Table 13. Incentive Amounts for Experiment 3

		Performance		
Test Venue	**Task**	**Good**	**Acceptable**	**Poor**
Test Track	Car Following	$1.20	$0.60	$0.0
	Secondary Task	$0.80	$0.40	$0.0
	Light-Detection	$0.40	$0.20	$0.0
	Total	$2.40	$1.20	$0.0

On each test track trial, the participant had the opportunity to earn $2.40. Thus, for good performance, each participant could earn an additional $24.00 in incentive pay.

The performance associated with each task in Experiment 3 was determined subjectively by the experimenter based on the general criteria presented in Table 14.

Table 14. Task Performance Incentive Criteria

Task	Good Performance	Acceptable Performance	Poor Performance
Car Following	Maintains close following distance consistently with minor deviations	Maintains close following distance mostly with some noticeable deviations	Generally fails to maintain close following distance
In-Vehicle Secondary Task	Performs secondary task continuously with minimal errors	Performs secondary task either intermittently or with moderate number of errors	Performs secondary task with considerable difficulty, slowly, and with moderate number of errors
Light-Detection	Consistently attentive to target detection, detecting most targets	Moderate number of targets not detected	Fails to detect significant number of targets

4.1.8 Data Reduction

Data were reduced to obtain the same measures that were used in STISIM in Experiments 1 and 2. These included: car-following coherence, car-following delay, headway, standard deviation of lane position (SDLP), steering entropy, PDT mean response time, PDT proportion of correct responses, head position X std, and RSME.

4.2 Results

We used Proc Mixed of SAS (Version 9.1.3) to compute an analysis of variance (ANOVA) for each dependent measure. Secondary task was the independent variable. It had the following nine levels:

1. Voice/Manual Destination Entry by Address
2. Baseline – no secondary task
3. Circles V1M1 – Easy visual discrimination, easy motor response
4. Circles V2M1 – Harder visual discrimination, easier motor response
5. Circles V2M2 – Harder visual discrimination, harder motor response
6. Voice/Manual Destination Entry by Place
7. Voice/Manual Selection of Previous Destination
8. Sternberg 3 – Easier memory task
9. Sternberg 7 – Harder memory task

As in Experiments 1 and 2, we planned comparisons to address specific questions concerning expected differences between the secondary task conditions. The comparisons, shown in Table 12, were adapted from those addressed in the first two experiments. Comparisons 1-4 were based on the reference/calibration tasks; comparisons 5-7 were based on the destination entry task.

Table 15. Planned Comparisons for Experiment 3

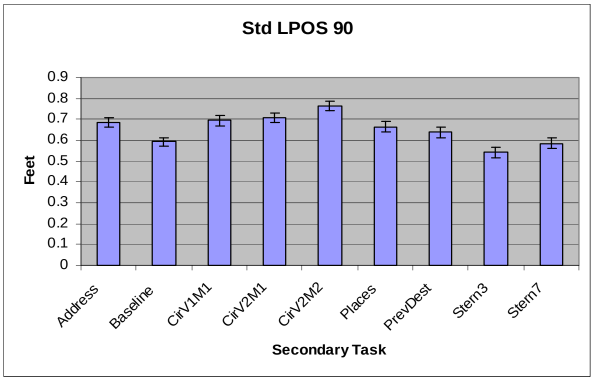
	Task	Comparison	Conditions
1	Circles	V1 vs. V2	3 vs. 4
2	Circles	M1 vs. M2	4 vs. 5
3	Circles	V1M1 vs. V2M2	3 vs. 5
4	Auditory/Vocal	Sternberg3 vs. Sternberg7	8 vs. 9
5	Destination Entry	Address vs. Previous Destination	1 vs. 7
6	Destination Entry	Point of Interest vs. Previous Destination	6 vs. 7
7	Destination Entry	Address + POI vs. Previous Destination	[1,6] vs. 7

Separate *F* tests were computed for each planned comparison. Probability values were adjusted for familywise error by using Hochberg's step-up method (Westfall et al., 2003). Adjusted *p* values of less than .05 are considered to be statistically significant; however, several tests with marginal results ($.05 < p < .10$) are considered noteworthy. A summary of the results of the planned comparisons with adjusted *p* values is presented in Table 16. Means for each performance measure by secondary task condition are presented in Figure 17 (pp. 46-47).

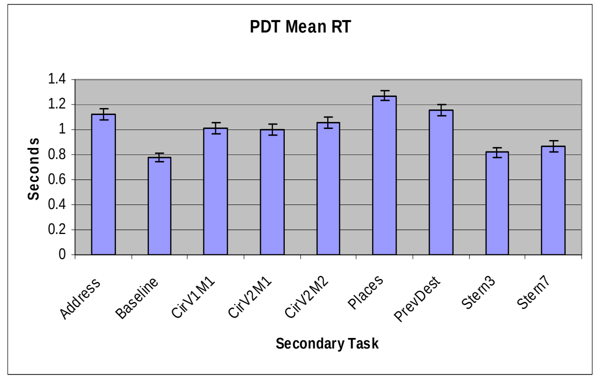
Table 16. Summary of Planned Comparison Results Experiment 3

	Task	Comparison	Delay	Cohere	Mean Hdwy	SDLP	Steer. Entropy	PDT MRT	PDT P Corr	Head X Std	RSME
1	Circle	V1 vs. V2	(.34)	(.98)	(.42)	(.68)	(.56)	(.76)	.03*	.006*	.0025*
2	Circle	M1 vs. M2	(.49)	(.98)	(.94)	(.16)	(.90)	(.63)	.09+	(.98)	(.12)
3	Circle	Both	(.49)	(.98)	(.42)	(.11)	(.24)	(.63)	.0004*	.015*	< .0001*
4	Auditory/Vocal	3 vs. 7	(.49)	(.99)	(.23)	(.21)	(.90)	(.63)	(.39)	(.98)	< .0001*
5	Navigation Destination Entry	Address vs. Previous Destination	.06+	.001*	(.15)	(.21)	(.90)	(.63)	.003*	(.98)	.0004*
6	Navigation Destination Entry	Point of Interest vs. Previous Destination	(.38)	(.36)	(.23)	(.43)	(.90)	(.34)	.0009*	(.98)	< .0001*
7	Navigation Destination Entry	Address + POI vs. Previous Destination	.09+	.009*	(.15)	(.21)	(.90)	(.63)	.0004*	(.98)	< .0001*

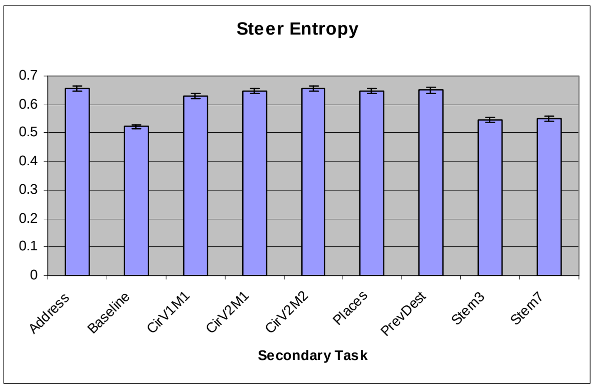
* Statistically significant difference (p < .05)
+ Marginally not significant (.05 < p < .10)
Parentheses denote differences that were not statistically significant



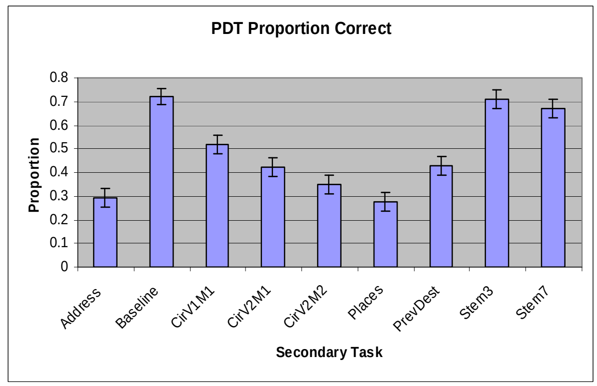
(a) Test Track SDLP



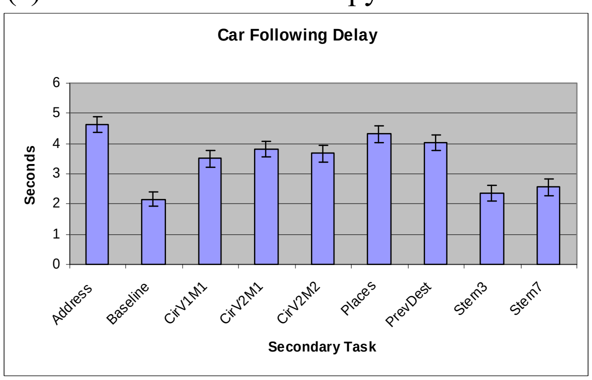
(b) Test Track PDT Mean RT



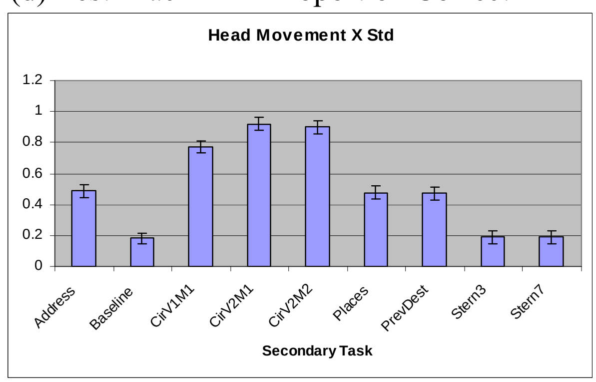
(c) Test Track Steer Entropy



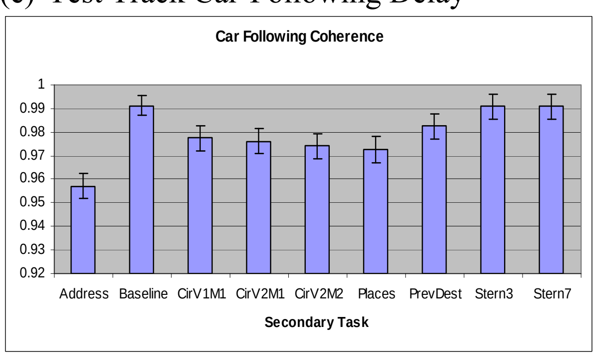
(d) Test Track PDT Proportion Correct



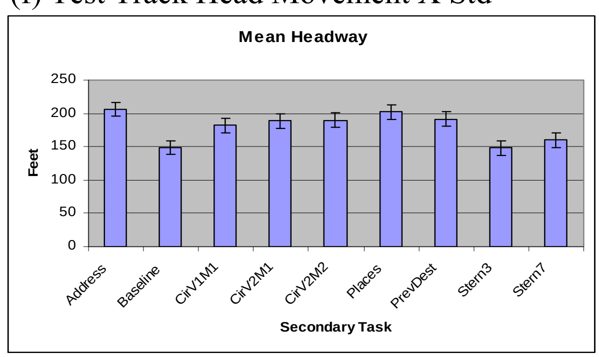
(e) Test Track Car-Following Delay



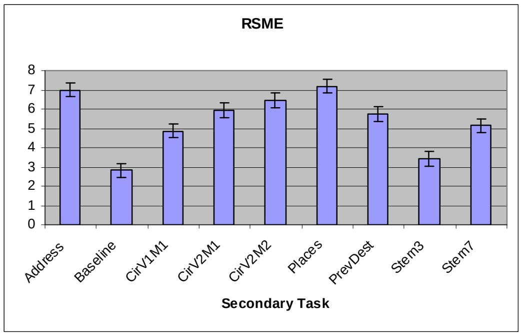
(f) Test Track Head Movement X Std



(g) Test Track Car-Following Coherence



(h) Test Track Mean Headway



(i) Test Track RSME (Workload Rating)

Figure 17. Mean values (± Standard Error) for Each Metric by Secondary Task Condition – Experiment 3

4.3 Discussion

As shown in Figure 17, baseline values were generally lowest for metrics reflecting performance degradation (e.g., SDLP, steering entropy, car-following coherence) and highest for metrics reflecting positive performance (e.g. car-following coherence, PDT proportion correct). This demonstrates the general sensitivity of the metrics to the secondary task loads used in this experiment.

As shown in Table 16 (PC 4), RSME was the only metric that was sensitive to the cognitive-only task differences between the two versions of the Sternberg memory-scanning task. Note also (in Figure 17) that for most metrics, the two Sternberg task means were closest to the baseline mean values. Thus, the objective metrics were generally not adequate for detecting impairment effects associated with this task. Whether the finding generalizes to other cognitive tasks is unknown, as is the question of how well this task simulates the demands of real-world secondary tasks.

In Experiment 3, the vehicle control and decision-making measures were generally not sensitive to the differences between secondary task conditions. The two car-following measures (coherence & delay) did exhibit sensitivity to two of the three navigation system comparisons (PCs 5 & 7); however, the other vehicle-based measures (SDLP, steering entropy, mean headway) did not detect any differences between secondary task conditions. This pattern of results was puzzling, since metrics used in this study, particularly SDLP, have demonstrated sensitivity for discriminating among different levels of these tasks in other experiments (e.g., Jamson & Merat, 2005). The failure of PDT Mean Response Time (MRT) to detect any differences was likely due to the relocation of the response button, noted in the discussion of Experiment 2.

Additional interpretation of results of Experiment 3, including comparisons with results from Experiments 1 and 2, is presented in the next section.

4.3.1 Comparison of Results from Experiments 1-3

In this section we compare means for selected measures across experiments. Comparisons are limited to the subsets of secondary tasks common to both lab and test track experiments. It should be noted also that because different participants were used in each experiment, some amount of unexplained variability due to group differences is expected. For this reason and because they were not included as planned comparisons, no statistical testing was done for comparisons of means across experiments.

RSME data from all three experiments are presented in Figure 18. RSME ratings were essentially identical for baseline driving in all three experiments, suggesting that the simulator experience generally replicated test track driving demands. RSME ratings were sensitive to differences between secondary tasks in all three experiments. Among Circles task conditions (CirV1M1, CirV2M1, CirV2M2), RSME ratings increased with increasing task demands in both Experiments 1 and 3. Circles task ratings were generally lower in Experiment 3 than in Experiment 1, which may reflect the fact that participants in Experiment 1 rated only the simple tasks, while those in Experiment 3 rated both simple and complex (navigation system) tasks. Thus, the Circles task may have seemed more demanding in the context of the simple tasks and less demanding in the context of the more complex navigation system tasks. RSME ratings differed between auditory/vocal task conditions (Stern3, Stern7) in both experiments. Among navigation system tasks, selecting a previous destination (PrevDest) was rated as less demanding than the two destination entry tasks (Address, Places) in both Experiments 2 and 3. The similarity among patterns of RSME ratings between laboratory and test track experiments implies that the simulator plus secondary task experience closely matched the test track plus secondary task experience.

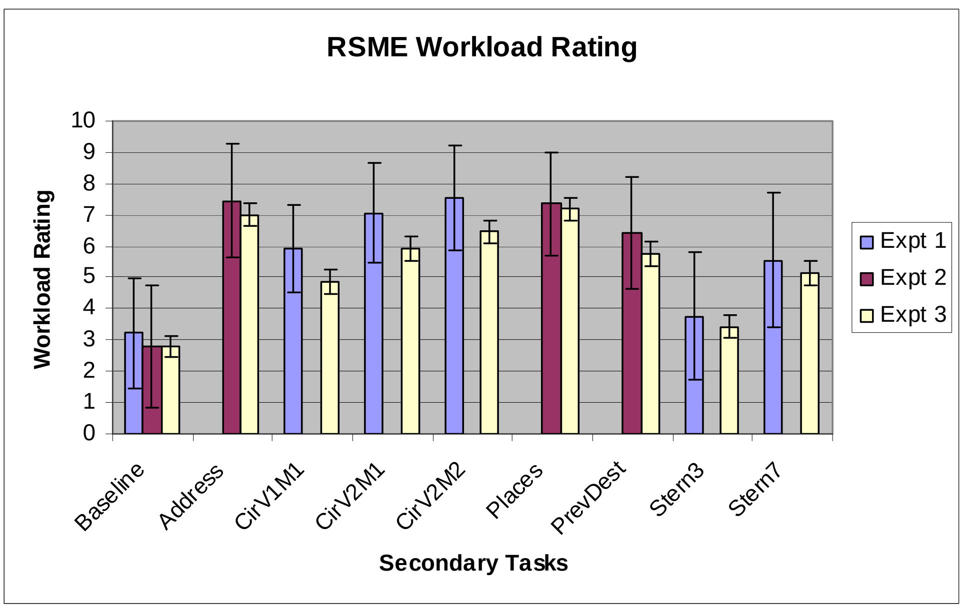


Figure 18. RSME Comparison Across Experiments 1 - 3

Summary results for PDT Mean Response Time are presented in Figure 19. The pattern of results was generally consistent across experiments. However, PDT mean response times were generally slower for Experiment 3 than for the comparable conditions in Experiments 1 and 2. This likely is due in part to target visibility differences between experiments. Specifically, PDT targets were often considerably more difficult to detect on the test track than in the simulator lab, particularly on sunny days. The change in PDT response button location is also a likely explanation for differences between Experiment 1 results and those for Experiments 2 and 3 observed on this measure. The response button was located on the driver's finger for Experiment 1, which allowed quick response without any movement or thought about the button location. In Experiments 2 and 3, the location was moved to the vehicle steering wheel to reduce the conflict between PDT responses and button presses necessary to activate the voice recognition system. The proximity of the two response buttons in the latter experiments may have created confusion. This, together with the added movement required, was probably responsible for the longer response times and decreased sensitivity of this measure in the latter two experiments. Consistent with this interpretation is the fact that mean differences between Experiments 1 and 3 (Circles and Stern conditions in Figure 19) are greater than those between Experiments 2 and 3 (Address, Places, PrevDest in Figure 19), in which the same button location was used.

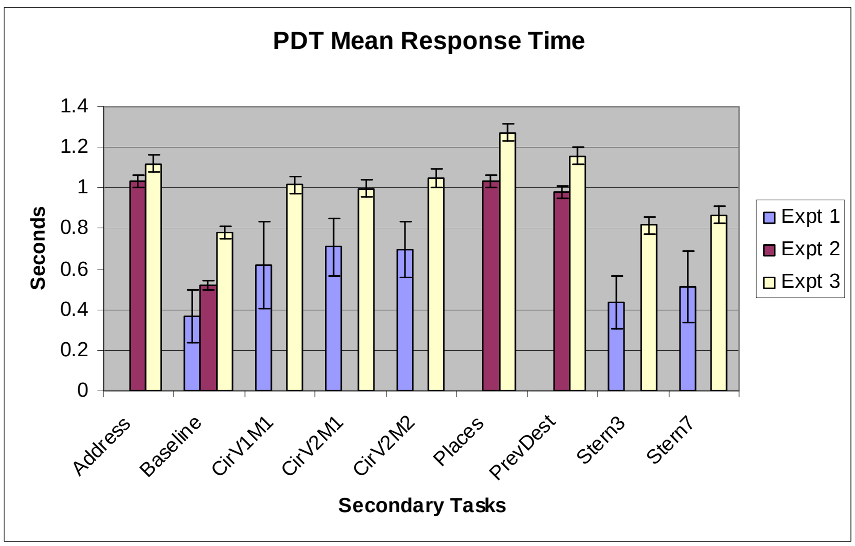


Figure 19. PDT Mean Response Time Comparison Across Experiments 1 - 3

Differences in target visibility were also the likely cause of consistent differences in the proportion of PDT targets detected that were observed between experiments. These data are shown in Figure 20. Drivers in Experiment 3 detected considerably fewer PDT targets than did drivers in the other experiments. Despite this difference, this measure was consistently sensitive to most differences between secondary task conditions, particularly in Experiments 2 and 3.

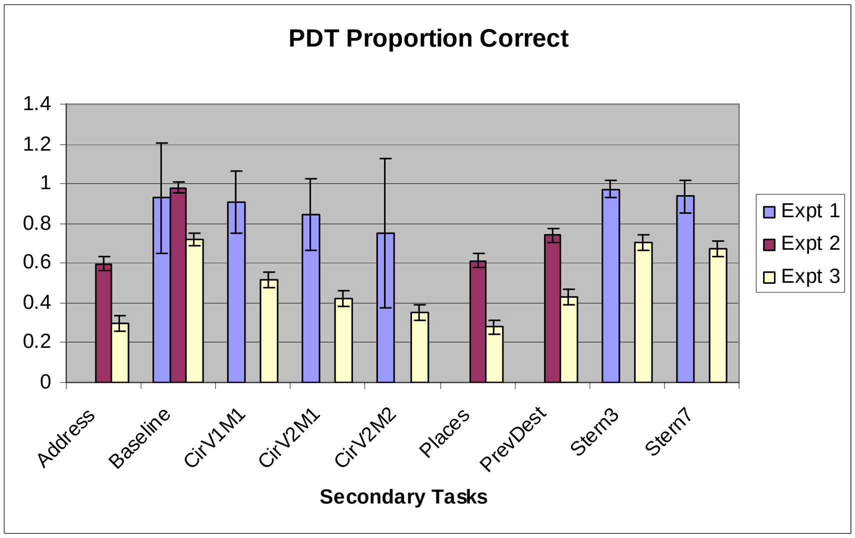


Figure 20. PDT Proportion of Correct Responses Comparison Across Experiments 1 - 3

The sensitivity of the standard deviation of lane position (SDLP) was not consistent across experiments. As shown in Figure 21, SDLP was highly sensitive to most of the manipulations in Experiments 1 and 2, but not in Experiment 3. There are several possible explanations for this difference. First, the lower sensitivity in Experiment 3 may be due to the presence of unexplained noise in the test track data, which necessitated differential treatment. Specifically, we cropped the data from Experiment 3 and computed summary measures using only 90 seconds, versus 180 seconds in Experiments 1 and 2. Second, SDLP is considerably easier to measure in simulators than in real driving situations. Simulators have direct information concerning the vehicle position in relation to the roadway at all times and have no difficulty computing this measure. In contrast, real-world computation of lane position depends on determining the lane boundaries, either with cameras or as in our study with a GPS test track survey. The test track measures may thus have included more error than the simulator measures of SDLP.

The SDLP data obtained from Experiment 3 revealed considerably less variability than was apparent in the two simulator experiments, which may have been related to steering system characteristics of the vehicle instrumented for test track use. Many newer vehicles have less steering system free play than older ones, which reduces the amount of lateral position variability that occurs when the drivers' hands are removed from the wheel for short periods of time. This possibility is supported by the finding that values of steering entropy, which measures steering error, were consistently smaller and less variable in Experiment 3 than in Experiments 1 and 2 (see Figure 22).

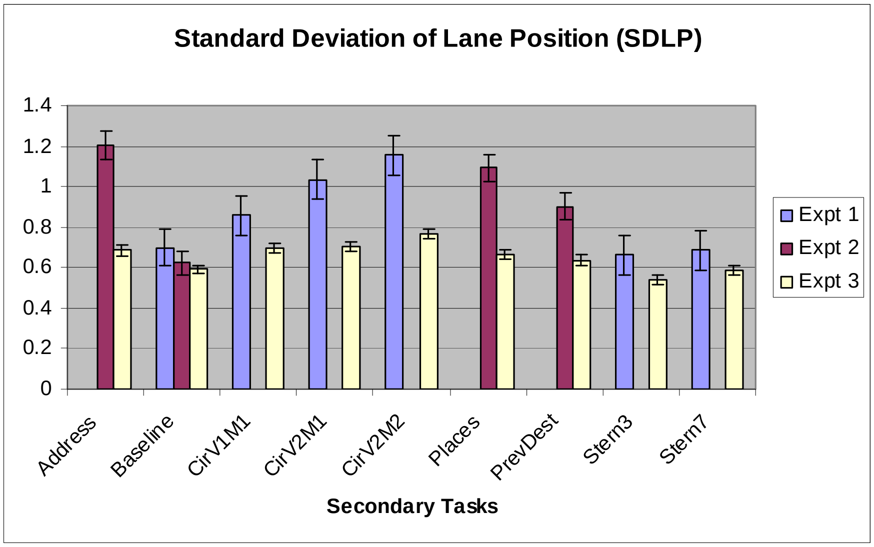


Figure 21. Standard Deviation of Lane Position (SDLP) Comparison Across Experiments 1 - 3

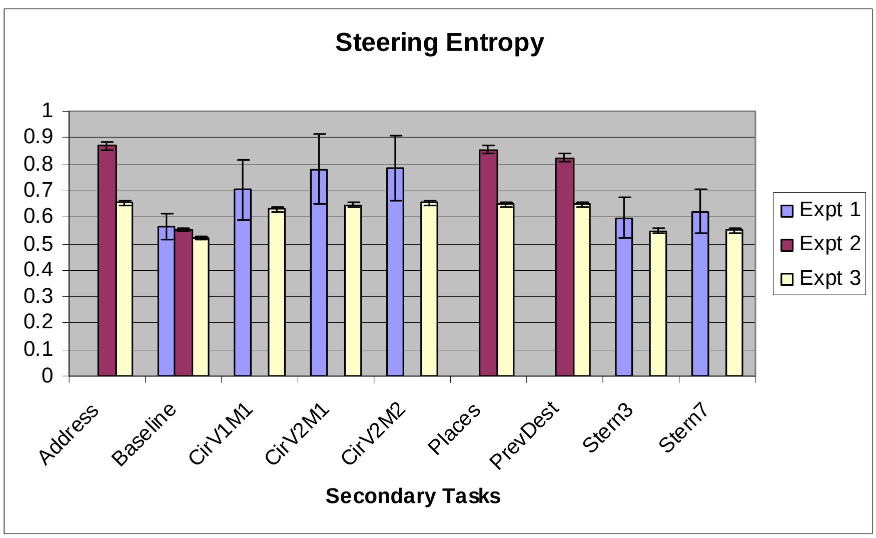


Figure 22. Steering Entropy Comparison Across Experiments 1 - 3

Due to the simulator's absence of motion cues and limited-fidelity of the visual display, we anticipated differences between the simulator and test track experiments among car-following measures. As shown in Figure 23, baseline car-following delay values were consistent across studies. Among the secondary tasks used in Experiment 3, only the Circles task results revealed consistently longer car-following delay values on the test track than in the laboratory. Differences for other secondary tasks (e.g., Address, Places, PrevDest, Stern3, and Stern7) were both smaller and mostly in the opposite direction, with shorter values occurring on the test track. This suggests that the Circles task was more disruptive to driving on the test track than in the laboratory. The most likely explanation is stimulus visibility differences, which made the Circles task more difficult during sunny conditions on the test track than in the laboratory.

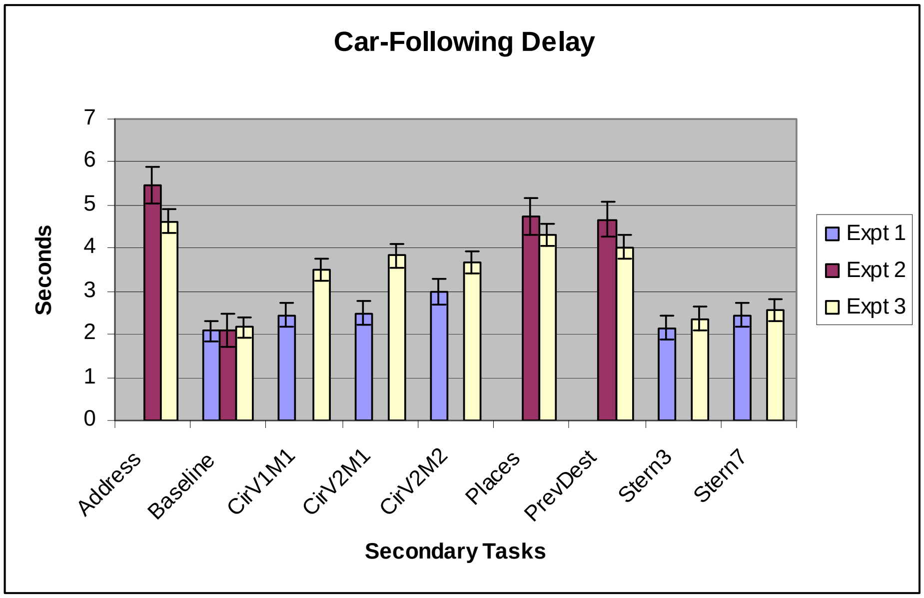


Figure 23. Car-Following Delay Comparison Across Experiments 1 – 3

Secondary task means for mean headway are presented in Figure 24. Drivers adopted longer headways in all conditions in the test track experiment. Again, there are several possible explanations for this difference. First, participants may have perceived the need to adopt longer following distances in the test track setting to compensate for the increased risk associated with driving a real vehicle relative to the lower risk associated with driving the simulator. Second, drivers in the test track study were not given active feedback about the following distance; such feedback was used in the laboratory study (see Section 2.1.5) because we were concerned that following distance would be more difficult to accurately judge in the simulator due to the limited fidelity of the visual display and the absence of motion cues.

If drivers did increase their headways to reduce driving task demands, we would expect to see a reduction in car-following performance. Typically, increased headways result in lower values for car-following coherence, which decreases the reliability of delay (phase shift) as a measure of response speed. In this study, however, coherence values remained extremely high, even with the longer headways observed in Experiment 3. These data are shown in Figure 25.

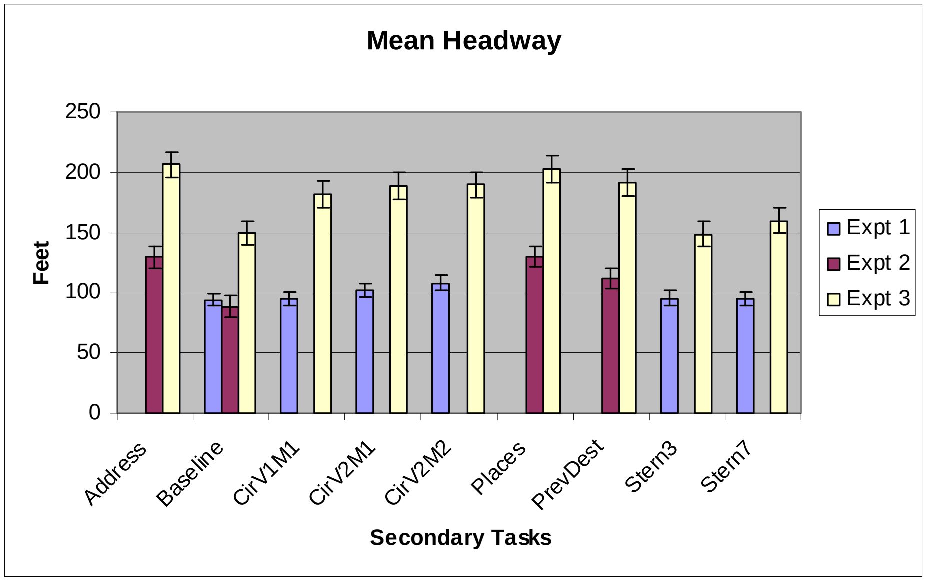


Figure 24. Mean Headway Comparison Across Experiments 1 - 3

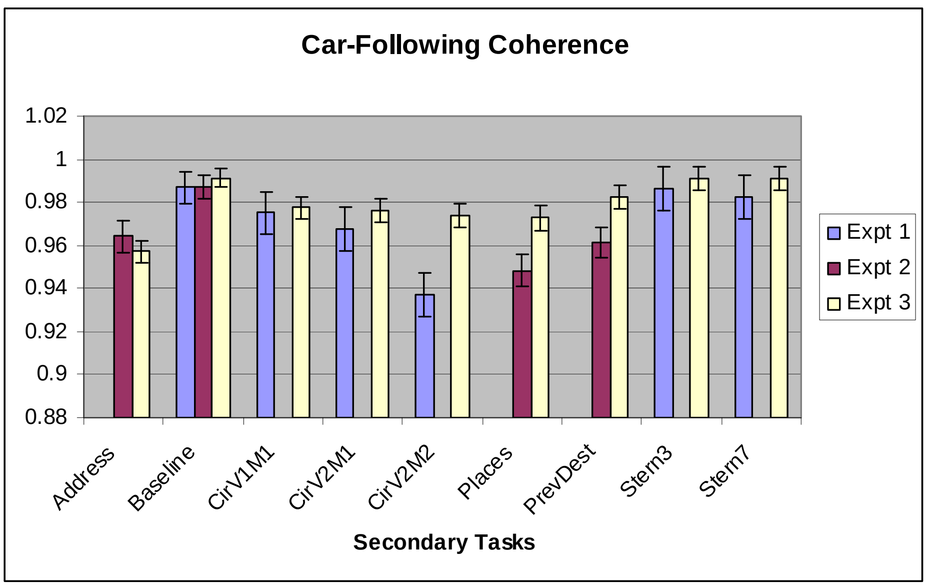


Figure 25. Car-Following Coherence Comparison Across Experiments 1 - 3

4.3.2 Secondary Task Comparisons

We compared secondary task performance data between lab and test track studies. The Circles task provided appropriate data for this purpose. This task was self-paced and participants were thus free to select the rate of secondary task completion. Mean times for first responses are presented in Figure 26.

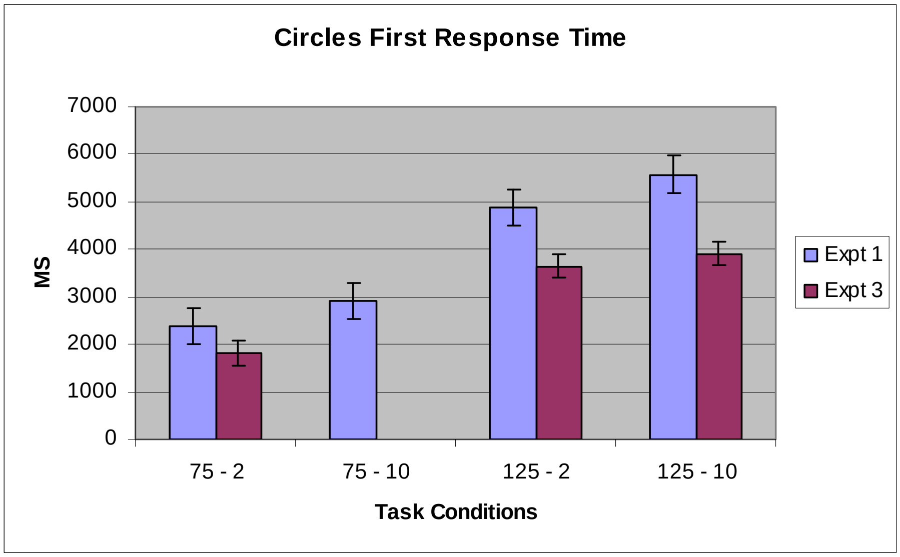


Figure 26. Circles Task First Response Time Across Experiments 1 and 3

Participants responded more quickly to Circles task trials in Experiment 3 than in Experiment 1. This was true across the three common conditions. A similar trend was observed for the trial response time, which represents the total time required for one trial. An increased number of secondary task trials completed per unit time in Experiment 3 is consistent with these findings.

Comparison of Circles task performance between Experiments 1 and 3 suggests that drivers devoted more attention to the Circles task in the test track study than in the laboratory study. Caution must be used in interpreting such differences directly because different participants were used in each experiment.

4.3.3 Simulator Validity

Our main objective with respect to simulator validity was to determine whether the simulator's limited fidelity and absence of motion cues reduced the sensitivity of simulator metrics relative to those obtained during test track driving. The results of Experiment 3 did not identify any metrics for which the test track sensitivity was greater than that found in the simulator experiments. Thus there was no loss of sensitivity in the simulator. However, this conclusion must be tempered by the fact that some test track metrics were less sensitive than their laboratory counterparts, primarily for detecting differences between conditions of the reference/calibration tasks used in Experiment 1. This pattern was apparent for SDLP and steering entropy, both measures of lateral vehicle control. Lateral position measures have consistently been found to be sensitive to differences in demand for visual/manual secondary tasks (Carsten & Brookhuis, 2005), such as the Arrows task used in Experiment 1 and the Circles task used in Experiments 1 and 3. If we were to consider the test track results as "ground truth," this raises the question of whether the simulator may be overly sensitive, such that some of the differences detected in the simulator experiments may not reflect real problems. However, the reference/calibration tasks have been used in numerous previous studies (e.g., Carsten & Brookhuis, 2005) and there is general agreement that the differences between conditions in these tasks represent meaningful differences in secondary task load. It should also be noted that we chose the Circles Task for

Experiment 3 because it represented more of a challenge for the metrics than the Arrows task, which was associated with consistently positive detection results for most metrics in Experiment 1. Therefore, we examined our test track methodology to identify problems that may have reduced the sensitivity of the associated metrics. As discussed above (Section 4.3.1), we identified a number of methodological and environmental considerations that may help explain the differences between the simulator and test track results. It is also possible that differences in drivers' risk perception between the simulator and test track venues, as evidenced by the consistently longer following distances adopted on the test track, may have reduced the sensitivity of the car-following measures on the test track; however, as indicated above the car-following results are not entirely consistent with this explanation. Additional experimental work will be necessary to evaluate these proposed explanations and until they can be resolved no definitive conclusions about the simulator validity can be made. However, we did not find deficiencies attributable to the reduced fidelity of the simulator driving experience. The main weakness of the simulator test venues is that neither did a good job of detecting differences between levels of cognitive distraction. The test track results were consistent with this finding. Finally, drivers' RSME ratings were consistent and equally sensitive across test venues, supporting the conclusion that the simulator and test track driving experiences were similar. This serves to establish face validity for the driving simulator.

5.0 DISCUSSION

In this section, we discuss the results of the three experiments within the context of developing a portable test to assess IVIS distraction potential. First, we consider the sensitivity of the metrics. Next, we identify the most promising metrics and identify several technical issues that need to be addressed as part of the test development. Finally, we compare the results of Experiment 3 with our previous experimental work using the same test track protocol.

5.1 Metric Sensitivity Based on Results of Experiments 1-3

Table 17 presents a summary of the results from all three experiments. Metrics have been categorized based on the number of differences successfully detected. There are several observations that can be made from this presentation. First, STISIM/PDT measures were generally more sensitive to the differences in secondary task load used in the present experiments than were LCT measures. Second, the highest sensitivity metrics were generally consistent across Experiments 1 and 2. Third, the highest sensitivity metrics were not consistent between the lab and test track studies. We now consider these observations in detail.

Among the objective measures, the STISIM measures of SDLP, Steering Entropy, PDT Mean RT (Experiment 1) and PDT Proportion Correct (Experiments 1-3) were most sensitive. The LCT summary measures were less sensitive as were the STISIM car-following measures.

RSME was most sensitive to the differences and was also consistent across all three experiments. As noted earlier, the sensitivity of RSME to all hypothesized differences in Experiment 1 led us to rely on it as an independent assessment of secondary task workload, particularly for navigation system tasks. At this point, one might ask why not just select RSME as the most promising measure of distraction potential. There are a number of well-known potential problems associated with the use of subjective rating scales (O'Donnell & Eggemeier, 1986; Gopher & Donchin, 1986). For example, subjective assessments depend on conscious experience and thus are limited to the task components of which the subject is aware (Gopher & Donchin, 1986). Aspects of cognitive distraction occurring outside of drivers' awareness may thus not be amenable to subjective assessment. Subjective assessments are typically retrospective and thus depend on the participants' working memory. Subjective assessments may confound the effects of mental and physical effort. They may also fail to distinguish between the external demands of a task and the effort invested to perform the task. Finally, numerous studies have found dissociation between subjective ratings and objective measures of task performance, which has raised concerns about relying on subjective measures alone (O'Donnell & Eggemeier, 1986). These concerns support our preference for objective measures in test development. Moreover, as test development proceeds the metrics will need to be tied more directly to safety. The objective metrics are more appropriate for this purpose.

The main purpose of Experiment 3 was to determine whether there were any driving task performance decrements observed in the test track experiment that were not also found in the laboratory experiments. Such findings would suggest that the laboratory simulator was not sufficiently sensitive for detecting all effects of secondary task load. The present results indicate that this was not the case. In fact, among the measures recorded both on the test track and in the (STISIM) simulator, the simulator measures were more sensitive to the differences between secondary task conditions than were the corresponding test track measures. Possible reasons, explained in greater detail in Section 4.3.1, include the increased difficulties of recording some measures in real-world versus laboratory settings, unwanted variability introduced by visibility problems due to sunny conditions, and the potential effects of newer vehicles' handling characteristics on vehicle control measures. The observed differences between lab and test track experiments need to be reconciled. However, with respect to the objectives of Experiment 3 the results lead us to conclude that the simulator's reduced fidelity was not associated with a loss of metric sensitivity.

5.2 STISIM/PDT versus LCT

To simplify decision making, product developers and policy makers desire a single metric that summarizes the distraction potential associated with different IVIS tasks. The major methodological studies (Angell et al., 2005; Carsten et al., 2005a) have concluded that the multidimensional nature of distraction is not consistent with the use of a single metric. The LCT represents an attempt to characterize three aspects of driving performance with a single metric. These include response time, target detection accuracy, and vehicle control. LCT developers have made provisions for computing their summary measures using both normative models and individual baselines. Some attempts have been made to decompose the LCT into its 3 components (Transport Canada); however, this approach has not had significant success. Despite the appeal of the LCT approach, neither LCT summary measure was as sensitive to the manipulations undertaken in Experiments 1 and 2 as the STISIM/PDT measures. The most promising STISIM/PDT metrics were SDLP, Steering Entropy, PDT Mean RT, and PDT Proportion Correct.

The STISIM/PDT offers the advantage not only of providing multiple measures that assess different aspects of driving performance, but also the flexibility for varying the demands of the driving task components. This capability may be important for addressing the several shortcomings of the STISIM/PDT test venue identified in this study. The most notable shortcoming was the inability of the metrics, with the exception of PDT Mean RT in Experiment 1, to differentiate between the two versions of the Sternberg memory scanning task, which was primarily a cognitive task. Additional work will be necessary to determine if changing the demands of the simulated driving task can improve the sensitivity of the STISIM/PDT metrics for detecting the effects of cognitive distraction.

Table 17. Comparison of Metric Sensitivity Across Experiments 1-3

Sensitivity	Experiment 1	Experiment 2	Experiment 3
High	STISIM: SDLP PDT Mean RT STISIM: Steer Entropy RSME	STISIM: SDLP STISIM: Steer Entropy PDT % Correct RSME	PDT % Correct RSME
Moderate	PDT % Correct LCT: Mean deviation STISIM: Car-following delay Head position variability	STISIM: Mean headway Head position variability LCT: Mean deviation LCT: Indiv. Baseline Mean deviation	Head position variability Car-following delay Car-following coherence
Low	LCT: Indiv. Baseline Mean deviation STISIM: Car-following coherence STISIM: Mean headway	PDT Mean RT STISIM: Car-following coherence STISIM: Car-following delay	SDLP Mean headway Steer Entropy PDT Mean RT

Three additional issues need to be considered in any future development. The first issue concerns the method of obtaining steering signals from drivers in stationary real vehicles. In this study, we used a separate steering wheel that was attached to the vehicle steering wheel. This created problems for evaluating systems that required drivers to press buttons located on the steering wheel. To the extent that IVIS may consistently require such button presses, reconsideration of a rotating plate option for recording steering inputs is warranted. The second issue concerns possible improvements to the PDT, which include reconfiguration of stimuli to cover a wider range of locations and the use of different sensory modalities. Issues related to the PDT have been addressed as part of the AIDE project (Merat et al., 2007). The third issue relates to difficulties we experienced with the faceLab eye tracking system. Many of the most successful metrics used in previous work, particularly the CAMP study, were based on eye glance data. In that study, the eye glance data were prepared manually, which is relatively labor intensive and probably not practical for routine use in a standardized test of distraction potential. We attempted, unsuccessfully, to create such metrics using an unobtrusive eye tracker. We have identified several technical improvements, including the capability of providing immediate feedback to the operator during calibration, which may improve system performance; however, unless reasonable quality eye tracker data can be obtained in a relatively time-efficient manner, we may need to conclude that metrics based on eye position or eye glance characteristics are not feasible for inclusion in a portable test of distraction potential.

Finally, once a set of test parameters has been defined, sufficient data will be needed to determine whether a single index of distraction potential can be developed by combining weighted values of a set of metrics.

5.3 Comparison with Previous Experimental Results

Previous research conducted at VRTC has used the same track protocol that was used in Experiment 3. In essence, drivers perform car following and peripheral target detection while engaged in potentially distracting secondary tasks. Two of our previous results have implications for the development of a test of distraction potential. The first concerns the detection of effects that are primarily cognitive. In our previous work (Ranney et al., 2007), we found effects of cognitive distraction using a simulated phone conversation. This task, designed by Baddeley (Baddeley, Logie, Nimmo-Smith, & Brereton, 1985), required drivers to listen to sentences, decide whether or not they made sense and also to remember either the subject or object of four sentences before reporting them aloud. This task was performed without any visual or manual components. Thus, like the Sternberg memory scanning task used in Experiments 1 and 3, it was primarily cognitive. In the earlier study, we found that steering entropy and both PDT measures were sensitive to the Baddeley task load and although not explicitly tested in the present study, it appears that differences between Sternberg task conditions and baseline were negligible. The results of the earlier study give us some confidence in the sensitivity of the metrics and at the same time raise the question of which cognitive task is more representative of the load of a real-world task that drivers might undertake while driving. In our experience, the Baddeley simulated phone conversation task is more demanding than a typical phone conversation. While both the Baddeley and Sternberg tasks require participants to keep items in memory, the Baddeley task is arguably more difficult because while remembering items, participants are also engaged in listening to and thinking about new material and making regular verbal responses. In contrast, the Sternberg task has a dead interval that requires only that participants mentally rehearse the items to maintain them in short-term memory. This suggests that our metrics are sensitive to cognitive loads but only if they are more demanding than those associated with the Sternberg memory-scanning task.

In our previous study, we also had drivers perform navigation tasks that required use of hierarchical menu systems, much like those used by the Acura navigation system. However, unlike the Acura system, these tasks did not require button presses or have a visual display. The metrics were sensitive to these effects and several metrics revealed higher levels of performance degradation (e.g., slower PDT response times) for the navigation tasks relative to the Baddeley phone conversation task. Thus, despite difficulties differentiating between Sternberg task levels in the present study, these previous results suggest that the metrics are sufficiently sensitive to effects of IVIS tasks, including those that are primarily cognitive.

The second issue raised by comparison with our previous work is based on the discrepant results provided by PDT measures. In our previous work, the two PDT measures (proportion correct, mean response time) had generally consistent results. Typically, secondary tasks were associated with longer response times and fewer targets detected. In the present study, only the proportion of targets detected remained reliable in detecting effects of most loads. We have tentatively concluded that this discrepancy is likely due to our decision to change the nature of PDT response in Experiments 2 and 3. In Experiment 1, we used the same response method as in the previous work and the results were generally consistent with previous results. In Experiments 2 and 3, we separated the response button from the driver and required the driver to move his/her hand each time a PDT target was detected. We hypothesize that the variability associated with hand movement time masked differences between conditions, thus reducing the sensitivity of the PDT response time measure. As part of future work, we need to compare performance between locations and, if warranted, return to the original method of recording button presses.

6.0 CONCLUSIONS

Based on the results of the present study, we conclude that:

1. Developing a test capable of assessing the distraction potential of IVIS in production vehicles appears to be feasible. Generally, the metrics considered in this work were more sensitive for detecting distraction effects associated with tasks with visual and manual demands than those with primarily cognitive demands. Additional work is necessary to determine whether the metrics' sensitivity for detecting effects of cognitive distraction can be improved. This is important in anticipation of the continued emergence of IVIS technologies with auditory/voice based interfaces.
2. Metrics obtained using the combination of STISIM (PC-based simulator) + Peripheral Detection Task (PDT) were more sensitive than the Lane Change Task (LCT) summary measures to the manipulations used in Experiments 1 and 2. Among the objective measures, lane position variability, steering entropy, and PDT response time were more sensitive than the LCT mean deviation scores. PDT response time was the only objective metric sensitive to the different levels of cognitive distraction used in Experiment 1.
3. The STISIM/PDT combination offers greater flexibility for fine-tuning scenario components and provides a wider range of performance measures than the LCT, which is a standardized test. This flexibility provides the potential for varying driving task demands in an attempt to increase metric sensitivity for detecting performance degradation due to cognitive distraction.
4. The eye tracker used in this work failed to provide data of sufficient quality to support computation of eye glance metrics based on eye position. Additional work is necessary to determine whether the quality of these metrics can be improved. In addition, for inclusion in a portable test of distraction potential, provisions must be made for reduction and analysis of eye tracker data in a time-efficient manner.
5. The comparison of simulator and test track results revealed that the simulator, with relatively inferior fidelity, was not associated with a significant loss of metric sensitivity relative to the test track metrics, which were derived in a situation that more closely resembles on-road driving. To the contrary, the simulator metrics revealed greater sensitivity for detecting differences among levels of widely used reference/calibration tasks. Additional testing would be useful to understand the causes of these differences and to provide a more complete validation of the simulator relative to test track driving.
6. Additional testing on a variety of vehicles is necessary to confirm applicability of the metrics to a range of IVIS tasks and the portability of the test across different vehicle types.

7.0 REFERENCES

Angell, L. S., Auflick, J. L., Austria, P. A., Kochhar, D. S., Tijerina, L., Biever, W. J., Diptiman, T., Hogsett, J. R., Jr., & Kiger, S. M. (2005). Driver Workload Metrics Project, Task 2 CAMP.

Baddeley, A., Logie, R., Nimmo-Smith, I., & Brereton, N. (1985). Components of Fluid Reading. Journal of Memory and Language, 24, 119-131.

Barickman, F. S. & Goodman, M. J. (1999). Micro DAS: In-vehicle portable data acquisition system. Transportation Research Record, 1689, 1-8.

Boer, E. R. (2000). Behavioral entropy as an index of workload. In (pp. 3-125-3-128).

Brookhuis, K. A., De Vries, G., & De Waard, D. (1991). The effects of mobile telephoning on driving performance. Accident Analysis and Prevention, 23, 309-316.

Brookhuis, K. A., Waard, D. d., & Mulder, B. (1994). Measuring driving performance by car-following in traffic. Ergonomics, 37, 427-434.

Burns, P. C., Trbovich, P. L., McCurdie, T., & Harbluk, J. L. (2005). Measuring distraction: Task duration and the lane-change test (LCT). In Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting (pp. 1980-1983). Santa Monica, CA: Human Factors and Ergonomics Society.

Carsten, O. & Brookhuis, K. (2005a). Issues arising from the HASTE experiments. Transportation Research Part F: Traffic Psychology and Behaviour, 8F, 191-196.

Carsten, O., Merat, N., Janssen, W. H., Johansson, E., Fowkes, M., & Brookhuis, K. (2005b). Human Machine Interaction and the Safety of Traffic in Europe (Rep. No. HASTE Final Report 1.0). University of Leeds: Institute for Transport Studies.

Gopher, D. & Donchin, E. (1986). Workload – An examination of the concept. In K.R. Boff, L. Kaufman, and J.P. Thomas (Eds.) Handbook of perception and human performance. Volume II: Cognitive Processes and Performance. New York: Wiley and Sons.

Harms, L. & Patten, C. (2003a). Peripheral detection as a measure of driver distraction. A study of memory-based versus system-based navigation in a built-up area. Transportation Research Part F, 6, 23-36.

Harms, L. & Patten, C. (2003b). Peripheral detection as a measure of driver distraction. A study of memory-based versus system-based navigation in a built-up area. Transportation Research Part F: Traffic Psychology and Behaviour, 6, 23-36.

ISO/TC 22, S. 1. Road vehicles – Ergonomic aspects of transport information and control systems — Simulated lane change test to assess in-vehicle secondary task demand. ISO/WD 26022, 1-31. 2004. International Organization for Standardization.
Ref Type: Generic

Jamson, A. H. & Merat, N. (2005). Surrogate in-vehicle information systems and driver behaviour: Effects of visual and cognitive load in simulated rural driving. Transportation Research Part F, F8, 79-96.

Mattes, S. (2003). The lane change task as a tool for driver distraction evaluation. In GfA/17th Annual Conference of the International-Society-for-Occupational-Ergonomics-and-Safety (ISOES) Stuttgart, Germany: Ergonomia Verlag OHG, Bruno-Jacoby-Weg 11, D-70597.

Merat, N., Johansson, E., Engstrom, J. A., Chin, E., Nathan, F., & Victor, T. W. (2007). Specification of a secondary task to be used in safety assessment of IVIS (Rep. No. IST-1-507674-IP). Adaptive Integrated Driver-Vehicle Interface.

O'Donnell, R.D. & Eggemeier, F.T. (1986). Workload assessment methodology. In K.R. Boff, L. Kaufman, and J.P. Thomas (Eds.) Handbook of perception and human performance. Volume II: Cognitive Processes and Performance. New York: Wiley and Sons.

Ranney, T. A., Harbluk, J. L., & Noy, Y. I. (2005). Effects of voice technology on test track driving performance: Implications for driver distraction. Human Factors, 47, 439-454.

Ranney, T. A., Mazzae, E., Baldwin, G. H. S., & Salaani, K. (2007). Characteristics of voice-based interfaces for in-vehicle systems and their effects on driving performance (Rep. No. DOT HS 810 867). Washigton, D.C.: U.S. Department of Transportation, National Highway Traffic Safety Administration.

Victor, T. W., Harbluk, J. L., & Engstrom, J. A. (2005). Sensitivity of eye-movement measures to in-vehicle task difficulty. Transportation Research Part F, 8F, 167-190.

Westfall, P. H., Tobias, R. D., Rom, D., Wolfinger, R. D., & Hochberg, Y. (2003). Multiple Comparisons and Multiple Tests Using SAS. Cary, North Carolina: SAS Institute Inc.

8.0 APPENDICES

8.1 Appendix A: Technical Modifications Necessary to Adapt Selected Protocols

Adapting the ADAM Lane Change Task (LCT) and the STISIM for use with stationary vehicles has necessitated the development of a method of obtaining simulator vehicle control inputs from the stationary vehicles. VRTC designed a steering wheel mechanism for use in a portable simulator application to be installed in subjects' personally owned vehicles (Figure 27). The design needed to be quickly installed, output high-resolution rotational data, and work with a large a variety of vehicles. It also should be able to interface and be easily calibrated within the STISIM Drive environment.

Although there are many designs that could accomplish this, it was determined that the most economical and timely method would be to install a secondary steering wheel on top of the OEM steering wheel (Figure 27(a)). Using this method the vehicle's steering system is not compromised and testing can be conducted without the vehicle running to provide steering boost.

The simulator steering wheel is installed on top of the real steering wheel. It is affixed with zip-ties or similar temporary clamps (Figure 27 (b)). When it is mounted, it axially extends the steering wheel approximately 3.5 inches. This distance could be shortened; however, enough room must be provided for the fingers so the participant may comfortably grab the wheel (Figure 27 (c)).

The steering wheel is coupled to a rotary bearing that is press fit into an aluminum sleeve. This bearing allows for smooth turning of the wheel even under conditions of heavy side loading. The shaft is coupled with a metal gear set that turns an optical encoder as the wheel moves. The optical encoder provides a quadrature output that is low in noise and high in resolution (Figure 27(d)).

Steering feel is accomplished by two methods. First, an elastic cord is used to provide a reverse torque as the subject turns the wheel in either direction past center (Figure 27 (e)). The cord provides a nonlinear torque that gets harder as the subject turns the wheel further away from 0 degrees. The cord limits the turning of the wheel to about 270 degrees in each direction. Second, a Delrin bushing is seated against the aluminum sleeve on the steering wheel shaft. The bushing can be adjusted to provide variable friction to the rotation of the wheel. The purpose of this is to provide damping to limit the amount of oscillation of the steering wheel on a quick return to center.

To properly use the generic steering wheel, the OEM wheel must stay in the locked position. In extreme cases, it is possible to overpower the OEM steering lock and cause inaccuracies in the data. If the driver is required to interact with an advanced vehicle system such as route navigation, the ignition may need to be "on" causing the OEM steering lock to disengage. An alternate method to stabilize the steering wheel will have to be used in these cases. Additionally, if any of the advanced systems require the driver to press a button on the steering wheel, these may be blocked or obscured.

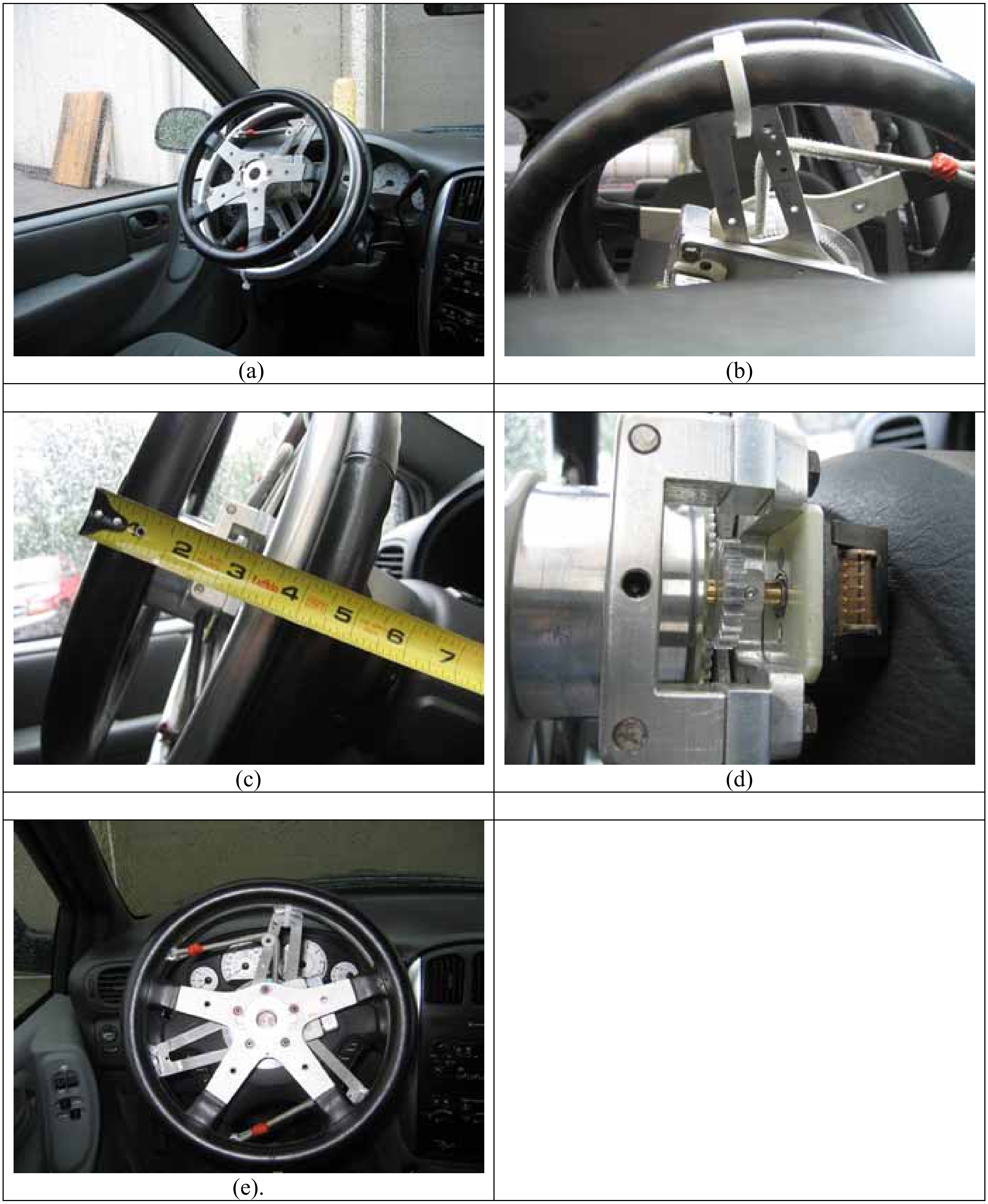


Figure 27. Steering Wheel Setup for Simulation.

8.2 Appendix B: Participant Information Summary for Simulator Protocols

STUDY: Attentional Demands of Operating In-Vehicle Devices
STERLING IRB ID: 2216
DATE OF IRB REVIEW: 03/29/06

PARTICIPANT INFORMATION SUMMARY AND CONFIDENTIAL INFORMATION FORM

STUDY TITLE: Attentional Demands of Operating In-Vehicle Devices: Establishing Best Practices

STUDY INVESTIGATOR: Thomas A. Ranney, Ph.D.

STUDY SITE: The Ohio State University Center for Automotive Research
930 Kinnear Road
Columbus, OH 43212

TELEPHONE: 800-262-8309

SPONSOR: National Highway Traffic Safety Administration

You are being asked to participate in a research study. Your participation in this research study is strictly voluntary, meaning that you may or may not choose to take part. To decide whether or not you want to be part of this research, the risks and possible benefits of the study are described in this form so that you can make an informed decision. This process is known as informed consent. This consent form describes the purpose, procedures, possible benefits and risks of the study. This form also explains how your information will be used and who may see it. You are being asked to take part in this study because the study investigator feels that you meet the qualifications of the study.

The study investigator or study staff will answer any questions you may have about this form or about the study. Please read this document carefully and do not hesitate to ask anything about this information. This form may contain words that you do not understand. Please ask the study investigator or study staff to explain the words or information that you do not understand. After reading the consent form, if you would like to participate, you will be asked to sign this form. You will be given a signed copy of your consent to take home and keep for your records.

PURPOSE

This research study is being conducted by the National Highway Traffic Safety Administration (NHTSA). The purpose of this study is to evaluate the different tools that researchers use to measure the level of distraction caused by "in-vehicle technologies." The latest in-vehicle technologies include devices that provide services such as access to the internet and navigation systems (for maps and driving directions), as well as the ability to send and receive e-mails. As new in-vehicle technologies are developed and marketed, there is a concern that these systems may interfere with driving. NHTSA is conducting this research study to determine the best way to collect data (information) on the use of in-vehicle technologies while driving.

Page 1 of 9 Participant's Initials__________________

STUDY REQUIREMENTS

You are being asked to participate in this research study because:

- You are 25 - 50 years of age,
- You have a valid, unrestricted U.S. driver's license (except for restrictions concerning corrective eyeglasses and contact lenses),
- You have a minimum of two years driving experience,
- You drive at least 7,000 miles per year, and
- You are in good general health.

NUMBER OF STUDY SITES AND STUDY PARTICIPANTS

This study will take place at one research site (The Ohio State University Center for Automotive Research) and will include at least 48 participants.

STUDY PROCEDURES

Before participating in this research study, you will be asked to read this Participant Informed Consent Form in its entirety. After all of your questions have been answered, you will be asked to sign this form to show that you voluntarily consent to participate in this research study.

Your participation in this research study will consist of one session lasting approximately 4 hours. During this session you will be asked complete specific driving objectives while performing different in-vehicle tasks. A member of the study staff will give you detailed instructions and will accompany you at all times during your participation in this research study.

Simulated Driving:

During your session you will be asked to drive a fixed-base simulator. A fixed-based simulator is a machine that imitates the conditions of driving in real life, but does not move. The simulator will be connected to the study vehicle, which will be an Acura TL (a mid-sized sport sedan). While driving the simulator you will sit in the driver's seat of the study vehicle. The study vehicle will have its engine turned off. You will control the simulator by moving an artificial steering wheel and the gas and brake pedals of the study vehicle.

The study vehicle will be equipped with sensors to collect information on your steering, braking and gas pedal usage. The sensors are located so that they will not affect your driving. The information collected by these sensors is recorded so that it can be analyzed at a later time. A large screen in front of the study vehicle will display a computer-generated image of the virtual road on which you will be driving.

Driving Objectives:

While operating the simulator, you will be asked to perform specific driving tasks. These tasks may involve activities such as following a car, changing lanes and/or detecting a light in your peripheral (side) vision.

 Participant's Initials___________________

In-Vehicle Tasks:

While completing the driving objectives you will be asked to perform specific in-vehicle tasks. These tasks will imitate or be similar to the actions required to operate in-vehicle technologies (such as a stereo, the internet, or a navigation system).

The in-vehicle tasks will consist either of tasks using a small computer screen located inside of the study vehicle or tasks using the stereo and navigation system in the study vehicle.

Eye Movement Recording and Monitoring:

Video cameras will be used to monitor your eye movements while operating the driving simulator and performing the in-vehicle tasks. The video cameras are located so that they will not affect your driving. The information collected using these video cameras is recorded so that it can be analyzed at a later time.

There are certain requirements for accurately recording your eye movements while driving. These requirements are as follows:

- Your entire face must be clearly visible while driving. If your hair hangs in your face, you may be asked to use clips or a rubber band to keep it out of your face.
- If you require corrective lenses and have contact lenses, you will be asked to wear them rather than glasses.
- You will not be permitted to wear sunglasses while driving.
- To help the eye tracking system better identify and track your facial features, you will be required to wear several small stickers on your face. The stickers will be put on before you begin driving and cannot be removed or moved until a member of the study staff informs you that you are finished driving. As a result you may be wearing the stickers for up to 3 hours.

Summary of Study Procedures:

The following procedures will take place at your session:

- After signing this consent form, you will be given instructions, training, and practice time for driving the simulator and performing the in-vehicle tasks.
- You will then complete a number of short tests, each lasting approximately 3 minutes. Each test will involve a different combination of driving objectives and in-vehicle tasks. You will be asked to complete approximately 30 tests (including all tests completed during training and practice).
- At the conclusion of the tests, you will be asked to answer brief questions about the tasks that you performed.
- After completing the questions the session will end, and your participation in this research study will be complete.

 Participant's Initials__________________

NEW INFORMATION

We do not anticipate that any changes to procedures will take place during this study. However, any new information developed during the course of the research that may affect your willingness to participate will be provided to you.

RISKS

Most people enjoy driving in the simulator and do not experience any discomfort. However, a small number of participants experience symptoms of discomfort associated with simulator disorientation. Previous studies with similar driving intensities and simulator setups have produced mild to moderate disorientation effects such as slight uneasiness, warmth, or eyestrain for a small number of participants. These effects typically last for only a short time, usually 10 -15 minutes, after leaving the simulator. If you ask to quit driving as a result of discomfort, you will be allowed to quit at once. You will be asked to sit and rest before leaving, while consuming a beverage and a snack. There is no evidence that driving ability is hampered in any way; therefore, if you show minimal or no signs of discomfort, you should be able to drive home. If you experience anything other than slight effects, transportation will be arranged through other means. This outcome is considered unlikely since studies in similar devices have shown only mild effects in recent investigations and evidence shows that symptoms decrease rapidly after simulator exposure is complete.

You will be asked to wear several small, latex stickers on your face while driving. These stickers may cause skin irritation in people with an allergy to latex. Allergic reactions may be mild (rash, hives) to severe (difficulty breathing, or a collapse of blood circulation and breathing systems). A severe allergic reaction, which is extremely unlikely, would require immediate medical treatment and could result in permanent disability or death.

There are no known physical or psychological risks associated with participation in this study beyond those described above.

BENEFITS

This research study will provide data on driver behavior and in-vehicle task performance that will be used by researchers to provide a scientific basis for developing recommendations or standards for performing in-vehicle tasks while driving. Your participation in this study will provide data that may help develop these recommendations or standards.

You are not expected to receive direct benefit from your participation in this research study.

ALTERNATIVES

This study is for research purposes only. Your alternative is to not participate.

CONDITIONS OF PARTICIPATION, WITHDRAWAL, AND TERMINATION

Participation in this research is voluntary. By agreeing to participate, you agree to operate the research vehicle in accordance with all instructions provided by the study staff. If you fail to follow instructions, or if you behave in a dangerous manner, you may be terminated from the study. At

Participant's Initials__________________

any time you may withdraw your consent and discontinue participation in the study at any time without penalty.

COSTS TO YOU

Other than the time you contribute, there will be no costs to you.

COMPENSATION

You will receive $20.00 per hour for the time you spend at the data collection facility. In addition, you will have the opportunity to earn incentive pay based on your performance on the driving and in-vehicle tasks. The maximum possible amount of incentive pay is $22.00.

If you voluntarily withdraw or are terminated from this study, you will be paid for the number of hours that you participated in the study.

USE OF INFORMATION COLLECTED

In the course of this study, the following data will be collected:

- Engineering data (such as the information recorded by the study vehicle sensors)
- Video/audio data (such as the information recorded by the video cameras)

Information NHTSA may release:

The **engineering data** collected and recorded in this study will include performance scores based on the data. This data will be analyzed along with data gathered from other participants. NHTSA may publicly release this data in final reports or other publications or media for scientific, educational, research, or outreach purposes.

The **video/audio data** recorded in this study includes your video-recorded likeness and all in-vehicle audio (including your voice). The video/audio data may include information regarding your driving performance. Video and in-vehicle audio will be used to examine your driving performance and other task performance while driving. NHTSA may publicly release video image data (in continuous video or still formats) and associated audio data, either separately or in association with the appropriate engineering data for scientific, educational, research, or outreach purposes.

Information NHTSA may not release:

Any release of **engineering data** or **video/audio data** shall not include release of your name. However, in the event of court action, NHTSA may not be able to prevent release of your name or other personal identifying information. NHTSA will not release any information collected regarding your health and driving record.

Participant's Initials__________________

QUESTIONS

Any questions you have about the study can be answered by Thomas Ranney, Ph.D., or the study staff by calling 1-800-262-8309.

If you have any questions regarding your rights as a research participant, you may contact Dr. Sally P. Green, Chairman of Sterling Institutional Review Board, 6300 Powers Ferry Road, Suite 600-351, Atlanta, Georgia 30339 (mailing address) at telephone number 1-888-636-1062 (toll free).

Participant's Initials___________________

INFORMED CONSENT

By signing the informed consent statement contained in this document, you agree that your participation is voluntary and that the terms of this agreement have been explained to you. Also by signing the informed consent statement, you agree to operate the study vehicle in accordance with all instructions provided by the study staff. You may withdraw your consent and discontinue participation in the study at any time without penalty.

NHTSA will retain a signed copy of this Informed Consent form. A copy of this form will also be provided to you.

Informed Consent Statement

I certify that:

- I have a valid, U.S. driver's license
- All personal and vehicle information as well as information regarding my normal daily driving habits provided by me to NHTSA, and/or the Transportation Research Center (TRC) employees associated with this study during the pre-participation phone interview and the introductory briefing was true and accurate to the best of my knowledge.
- I have been informed about the study in which I am about to participate.
- I have been told how much time and compensation is involved.
- I have been told that the purpose of this study is to evaluate the tools that researchers use to measure driving and in-vehicle task performance.
- I agree to operate the research vehicle in accordance with all instructions provided to me by the study staff.

I have been told that:

- The study will be conducted on a fixed-base driving simulator and that the risk of discomfort associated with simulator disorientation is minimal.
- For scientific, educational, research, or outreach purposes, video images of my driving which will contain views of my face and accompanying audio data may be used or disclosed by NHTSA, but my name and any health data or driving record information will not be used or disclosed by NHTSA.
- My participation is voluntary and I may refuse to participate or withdraw my consent and stop taking part at any time without penalty or loss of benefits to which I may be entitled.
- I have the right to ask questions at any time and that I may contact the study investigator, Thomas Ranney, Ph.D., or the study staff at 937-666-4511 or 800-262-8309 for information about the study and my rights.

Participant's Initials__________________

I have been given adequate time to read this informed consent form. I hereby consent to take part in this research study.

I,____________________________________, voluntarily consent to participate.
(Printed Name of Participant)

____
Signature of Participant   Date

INFORMATION DISCLOSURE

By signing the information disclosure statement contained in this document, you agree that the National Highway Traffic Safety Administration (NHTSA) and its authorized contractors and agents will have the right to use the NHTSA engineering data and the NHTSA video data for scientific, educational, research, or outreach purposes, including dissemination or publication of your likeness in video or still photo format, but that neither NHTSA nor its authorized contractors or agents shall release your name; and you have been told that, in the event of court action, NHTSA may not be able to prevent release of your name or other personal identifying information. NHTSA will not release any information collected regarding your health and driving record, either by questionnaire or medical examination. Your permission to disclose this information will not expire on a specific date.

Information Disclosure Statement

I, ______________________________________, grant permission to the
(Printed Name of Participant)
National Highway Traffic Safety Administration (NHTSA) to use, publish, or otherwise disseminate NHTSA engineering data and NHTSA video image data, as defined in this Participant Informed Consent Form (including continuous video and still photo formats derived from the video recording), and associated in-vehicle audio data collected about me in this study, either separately or in association with the appropriate engineering data for scientific, educational, research, or outreach purposes. I have been told that such use may involve widespread distribution to the public and may involve dissemination of my likeness in video or still photo formats, but will not result in release of my name or other identifying personal information by NHTSA or its authorized contractors or agents. I have been told that my permission to disclose this information will not expire on a specific date.

____
Signature of Participant Date

 Participant's Initials________________

8.3 Appendix C: Rating Scale Mental Effort (RSME)

Instructions

We are interested not only in assessing your performance but also the experiences you will have during the different task conditions. Right now I will describe the technique that will be used to examine your experiences.

Most importantly, we want to assess the mental effort you experience. Mental effort is a difficult concept to define precisely, but a simple one to understand generally. The factors that influence your experience of mental effort may come from the task itself, your feelings about your own performance, how much effort you put in, or the stress and frustration you felt. The mental effort contributed by different task elements may change as you get more familiar with a task, perform easier or harder versions of it, or move from one task to another.

Since mental effort is something experienced individually by each person, there are no effective "rules" that can be used to estimate the mental effort of different activities. One way to find out about mental effort is to ask people to describe the feelings they experienced. We will be using a rating scale to assess your mental effort. Please read the definition of the scale carefully. If you have a question about the scale, please ask me about it. It is extremely important that it is clear to you. The description will be made available to you for reference during the experiment.

Rating Scale Definition
Mental Effort: How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving? How hard did you have to work mentally? How much time pressure did you feel?

After performing a set of tasks, you will be instructed to bring the vehicle to a stop at a specified location. While the vehicle is stopped, the rating scale will be presented to you. You will evaluate the tasks performed (some combination of car following, light detection and phone tasks) since the time when the previous rating scale was administered, by telling the in-vehicle experimenter the number on the scale at the point that matches your experience. Please consider your responses carefully in distinguishing among the different task conditions. Your ratings will play an important role in the evaluation being conducted, thus your active participation is essential to the success of this experiment, and is greatly appreciated.

Subject: Date:

Condition: Exp:

Rating Scale Mental Effort

Modified 8/25/03

Please indicate, by marking the scale below, how much effort it took for you to perform the task you just completed

10	Extreme Effort
9	
8	Great Effort
7	
6	Moderate Effort
5	
4	Some Effort
3	
2	Little Effort
1	
0	Absolutely No Effort

Rating Scale Definition
Mental Effort: How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving? How hard did you have to work mentally? How much time pressure did you feel?

8.4 Appendix D: Simulator Sickness Questionnaire

Directions: Circle one option for each symptom to indicate whether that symptom applies to you right now.

1. General Discomfort.................... None................ Slight............... Moderate.......... Severe
2. Fatigue .... None................ Slight............... Moderate.......... Severe
3. Headache .... None................ Slight............... Moderate.......... Severe
4. Eye Strain .... None................ Slight............... Moderate.......... Severe
5. Difficulty Focusing .... None................ Slight............... Moderate.......... Severe
6. Salivation Increased .... None................ Slight............... Moderate.......... Severe
7. Sweating .... None................ Slight............... Moderate.......... Severe
8. Nausea .... None................ Slight............... Moderate.......... Severe
9. Difficulty Concentrating .... None................ Slight............... Moderate.......... Severe
10. "Fullness of the Head" .... None................ Slight............... Moderate.......... Severe
11. Blurred Vision .... None................ Slight............... Moderate.......... Severe
12. Dizziness with Eyes Open .... None................ Slight............... Moderate.......... Severe
13. Dizziness with Eyes Closed ...... None................ Slight............... Moderate.......... Severe
14. *Vertigo .... None................ Slight............... Moderate.......... Severe
15. **Stomach Awareness .... None................ Slight............... Moderate.......... Severe
16. Burping .... No .... Yes .... If yes, no. of times ________
17. Vomiting .... No .... Yes .... If yes, no. of times ________
18. Other ____________________________________

* Vertigo is experienced as loss of orientation with respect to vertical upright.

** Stomach awareness is usually used to indicate a feeling of discomfort which is just short of nausea.

Figure 28. Simulator Sickness Questionnaire

8.5 Appendix E: TRC Policy for Safeguarding Proprietary Information

Transportation Research Center Inc.

POLICY & PROCEDURE

CONFIDENTIAL INFORMATION		P&P NO.	153
Volume:	I, General Information	Issue Date:	11/30/2005
Function:	Security	Effective Date:	11/30/2005
Replaces:	Safeguarding Proprietary Info Issued 10/20/03	Code:	B, D

1. Purpose

To establish standards for the protection of confidential information and a proprietary atmosphere for TRC Inc. and its customers.

2. Scope

This policies applies to all customers and other visitors who have access to testing or other confidential information.

3. Policy

It is the policy of TRC Inc. to protect the identity, objectives, and presence of our customers, their test results, and/or other confidential information by the enforcement of the rules that are outlined herein. These rules are applicable to all personnel at/or within the facilities of TRC Inc.

3.1 You will not be allowed to witness any test or access other confidential information that you are not directly associated with unless prior approval has been given by facility management. This same restriction applies to the photographing of any test or test article.

3.2 In any activity that you are not directly associated with that you do witness, you agree not to disclose any information that you may have obtained.

3.3 Any violation of this policy may result in censure by TRC Inc. and possible punitive legal action through the courts.

I have read and understand the above P&P #153, Confidential Information, and accept my responsibilities in complying with this policy.

____________________________ ____________________________

Printed Name Signature

____

Company Name

____________________________ ____________________________

Witness Signature Date

8.6 Appendix F: Data Collection and Computational Methods for Deriving Performance Measures

This section will discuss data collection and the calculation of measures used to evaluate performance of car following. The performance measures include: mean and standard deviation of lane position, mean and standard deviation of headway, speed coherence, phase delay of speed, and speed modulus.

Sinusoidal speed profile

The lead vehicle speed is a sinusoid of 7.5 mph amplitude, +55 mph offset, and a one cycle per 30 seconds (0.033 Hz) frequency. Figure 29 shows the lead vehicle speed profile together with the resulting subject vehicle speed for a typical trial.

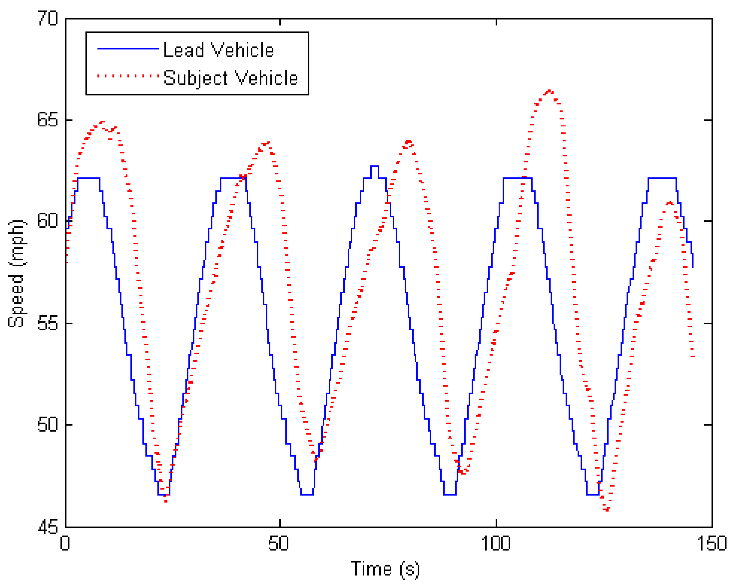


Figure 29. Sinusoidal speed profiles of lead and subject vehicles for a typical trial

Vehicle position and speed data collection

The lead vehicle speed is obtained via laser speed sensor recorded on the MicroDAS at 30-Hz sampling rate. The lead vehicle MicroDAS collects lead vehicle data for all trials of any subject in a single MicroDAS file.

The subject vehicle speed and position are obtained via a Global Positioning System (GPS) receiver and a devoted in-vehicle laptop. The laptop stores GPS speed and position data at 10-Hz sample rate in a .gpb file using GPS Data Logger software from Waypoint Consulting. An event switch starts/stops lead vehicle MicroDAS collection and simultaneously triggers a one-shot circuit that marks start and end times of the event in a Waypoint GPS station file. Thus, one subject vehicle MicroDAS file, one event start GPS time stamp, and one event stop GPS time stamp are recorded per trial.

Similarly, Waypoint GPS base station data is collected from a stationary receiver and PC running Waypoint Data Logger. The combined base station and lead vehicle Waypoint .gpb files are used to process differential GPS. The Waypoint GPS software is capable of outputting position and velocity data at 10 Hz to locally stored files (a base station file and a subject vehicle file). The two Waypoint GPS files are processed using Waypoint GrafNav software (combined forward and reverse solution) and time, position, and velocity data are exported to a GPS text file.

GPS Sync, an in-house developed program, merges the Waypoint GPS text file and Waypoint station mark file with the MicroDAS files from both vehicles. Because Waypoint GPS data is 10 Hz and MicroDAS data is 30 Hz, GPS Sync linearly interpolates the subject vehicle Waypoint GPS data. The lead vehicle laser speed data is retained in original sample-and-hold format (refer to Figure 29).

Loss of subject vehicle GPS signal

The test track has two bridge overpasses which cause dropouts in the Waypoint GPS signal, thus resulting in momentary loss of subject vehicle position and speed data. The length of dropout can be approximately 110 to 125 meters depending on vehicle speed at the time of passing underneath an overpass.

GPS Sync linearly interpolates missing segments of position and speed data. In some instances, GPS Sync subject speed interpolation needs to be corrected due to Waypoint GrafNav processed speed variances just prior to dropout. In all cases, Matlab is employed to extend the beginning of subject vehicle speed interpolation several samples prior to the dropout. This correction was not necessary for subject vehicle GPS position data. Figure 30 shows an example of a Matlab corrected interpolation of a speed dropout.

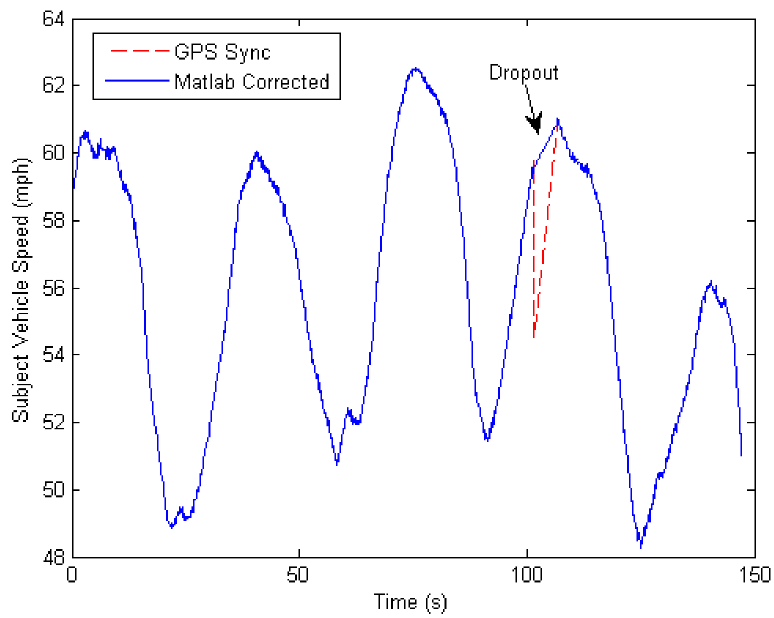


Figure 30. Correction of subject vehicle speed interpolation

Subject vehicle lane position

A combination of Waypoint GPS and lane tracking were previously used to obtain a detailed survey of the lane markings of the high-speed test track. Since the GPS antenna is installed at the centerline of the subject vehicle, every GPS position point collected represents a "distance from line" measure when perpendicular distance to the line is calculated. Because both lanes 1 and 2 were utilized, lane position is determined with respect to the line dividing lanes 1 and 2. Measures to the left of the dividing line (subject driving in lane 2) are negative, while measures to the right of the dividing line (subject driving in lane 1) are positive.

Lead vehicle to subject vehicle headway

An Eaton VORAD system is installed on the front of the subject vehicle in order to provide forward-looking radar range data. For redundancy, the same VORAD system is installed on the rear of the lead vehicle to provide rearward-looking radar range data. The range data is recorded at 30 Hz on the respective vehicle's MICRODAS.

Frequency Analysis

The lead vehicle speed profile has a fundamental frequency of 0.033 Hz. The subject vehicle speed performance is evaluated with respect to this fundamental frequency since other values are contaminated with spectral leakage.

Matlab's Transfer Function Estimate function (TFESTIMATE) is used to estimate the transfer function of the lead vehicle speed/subject vehicle speed system. The magnitude of this transfer function results in the system modulus. The phase angle of this transfer function results in the phase delay. The format of the function is:

TFESTIMATE(X,Y,WINDOW,NOVERLAP,NFFT,Fs)

Matlab's Power Spectral Density (PSD) function (PWELCH) is used to estimate PSD of lead vehicle and subject vehicle speeds individually using Welch's method. The frequency at which maximum subject vehicle PSD occurs is the evaluation frequency for coherence. The format of the PWELCH function is:

PWELCH(X,WINDOW,NOVERLAP,NFFT,Fs)

Matlab's Magnitude Squared Coherence Estimate function (MSCOHERE) is used to calculate coherence directly. Coherence is evaluated at the frequency of maximum PSD. The format of the function is:

MSCOHERE(X,Y,WINDOW,NOVERLAP,NFFT,Fs)

X is the input signal. For TFESTIMATE and MSCOHERE, X is the lead vehicle speed signal, normalized by subtracting the mean of the lead vehicle speed. For PWELCH, PSD is estimated twice using both normalized lead vehicle speed and normalized subject vehicle speed as input X.

Y is the output signal. This is the subject vehicle speed signal, also normalized by subtracting the mean lead vehicle speed. NFFT, the number of Fast Fourier Transform points, is one-half the length of the subject speed signal.

WINDOW is an NFFT-point Hann (Hanning) window. A Hann window is used because of its ability to reduce 'leakage' of power around the main lobe. Without use of an appropriate window, the spectral estimates can be distorted. Since the Hann window discards relevant information at the beginning and end of each record, the ensemble-averaged records are overlapped by 80% of window duration.

NOVERLAP, number of overlap samples, is 80% of NFFT (rounded toward zero to nearest integer).

Fs is the sampling frequency of 30 Hz.

Figure 31 shows the magnitude and phase plots of the transfer function resulting from the typical trial shown in Figure 29.

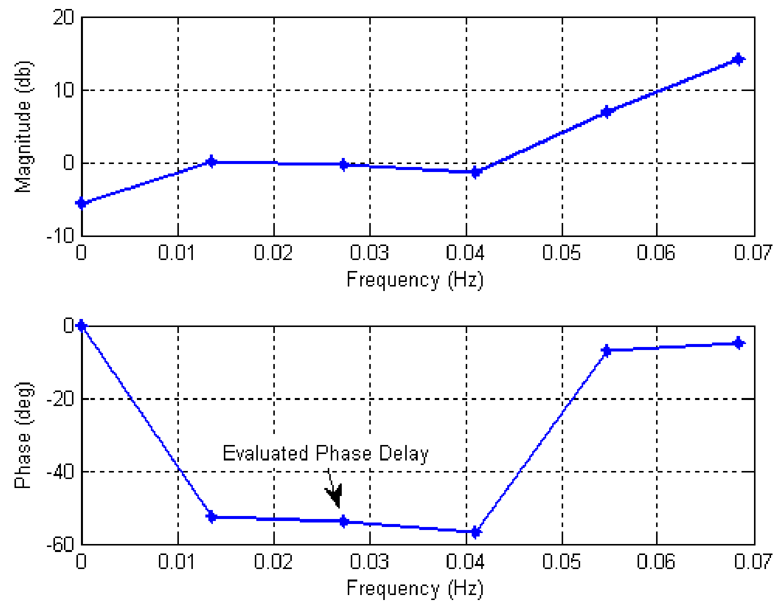


Figure 31. Magnitude and phase delay of lead vehicle/subject vehicle transfer function

Figure 32 shows the normalized lead vehicle and subject vehicle PSD for the typical trial shown in Figure 31. While the lead vehicle speed profile is a 0.033 Hz sinusoid, the PSD plot does show that the power spectrum around this frequency exhibits spectral leakage. Using measures at 0.033Hz is valid and there is no need for arithmetic weighting of multiple points.

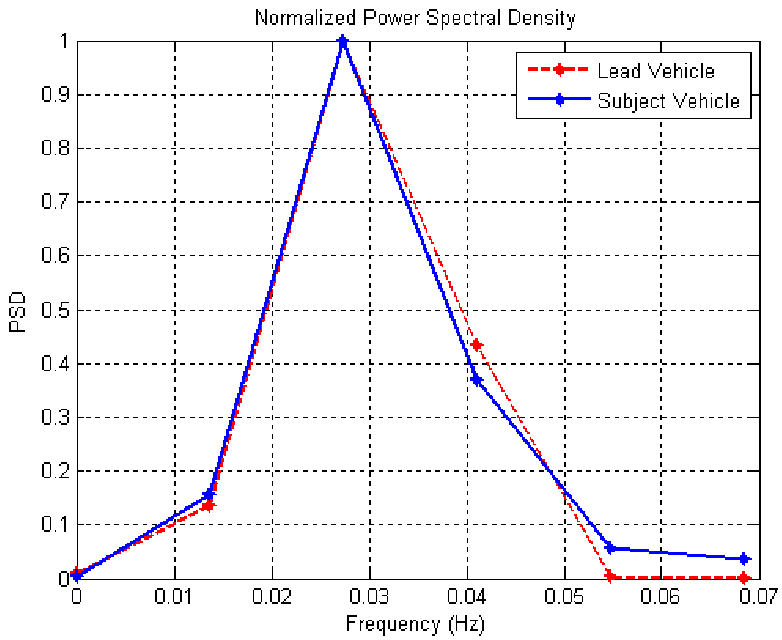


Figure 32. Normalized power spectral density of lead and subject vehicle speeds

Figure 33 shows the coherence plot for the typical trial shown in Figure 29.

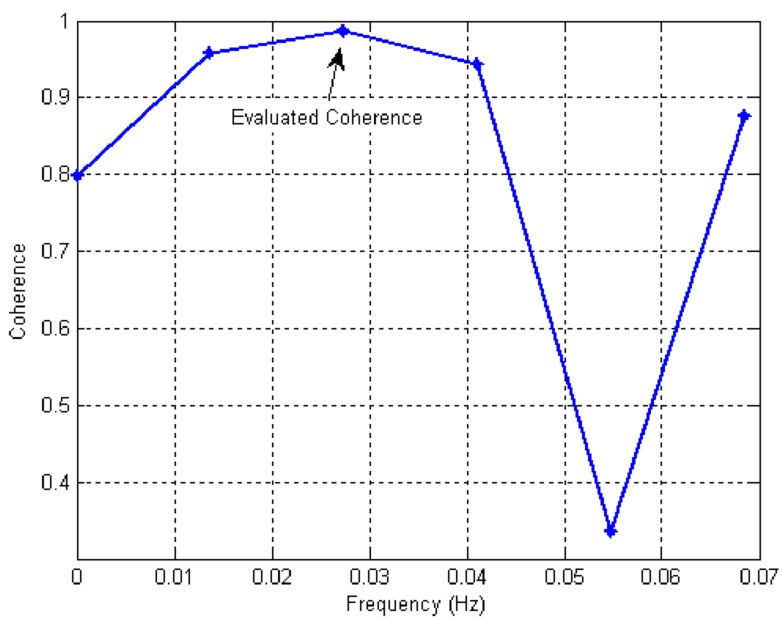


Figure 33. Speed coherence of subject vehicle with respect to lead vehicle

The following Matlab code was used to obtain phase delay, coherence, and modulus.

Calculation of Phase Delay, Coherence, and Modulus in Matlab

```
% SpdBLLc = Lead vehicle speed in mph
% SpdBLSc = Subject vehicle speed in mph
ns=1;  %set number of subjects to 1
Fs = 30;  %Sampling frequency
nfft = ceil(length(SpdBLSc)/2);
%
[coh0 F] = mscohere(SpdBLSc-mean(SpdBLSc),SpdBLLc-
mean(SpdBLLc),hann(nfft),fix(nfft*80/100),nfft,Fs);

[tf F] = tfestimate(SpdBLLc-mean(SpdBLLc),SpdBLSc-
mean(SpdBLSc),hann(nfft),fix(nfft*80/100),nfft,Fs);
%
[PsdS F] = pwelch(SpdBLSc-mean(SpdBLSc),hann(nfft),fix(nfft*80/100),nfft,Fs);
[Psdl F] = pwelch(SpdBLLc-mean(SpdBLLc),hann(nfft),fix(nfft*80/100),nfft,Fs);
%
is = find(PsdS == max(PsdS));
il = find(Psdl == max(Psdl));
mag = abs(tf);
phase = unwrap(angle(tf))*180/pi;
%
Subject(ns).Sine.BL.Coh = coh0(is);
Subject(ns).Sine.BL.Mag = mag(is);
if phase(is) > 0
    phase = phase - 360;
end
Subject(ns).Sine.BL.Pha = phase(is)/F(is)/360;
Subject(ns).Sine.BL.Frq = F(is);
%
tfang=unwrap(angle(tf(is)))*180/pi-360;  %transfer function angle at
fundamental frequency
if tfang<=-360  %
    num360=fix(tfang/360);
    tfang=tfang-num360*360;
end
% Delay, coherence, and modulus
phasedelay=-tfang/.0333/360;  %Delay -- 0.0333 is the fund freq (lead vehicle
speed)
fundcoherence=Subject(ns).Sine.BL.Coh;  %Coheherence
modulusfft=sqrt(real(tf(is))^2+imag(tf(is))^2);  %Modulus
```

DOT HS 811 231
December 2009

U.S. Department of Transportation
National Highway Traffic Safety Administration

www.nhtsa.gov